# Genomics and the origin of species

Seehausen, Ole ; Butlin, Roger K ; Keller, Irene ; Wagner, Catherine E ; Boughman, Janette W ; Hohenlohe, Paul A ; Peichel, Catherine L ; Saetre, Glenn-Peter ; Bank, Claudia ; Brännström, Ake ; Brelsford, Alan ; Clarkson, Chris S ; Eroukhmanoff, Fabrice ; Feder, Jeffrey L ; Fischer, Martin C ; Foote, Andrew D ; Franchini, Paolo ; Jiggins, Chris D ; Jones, Felicity C ; Lindholm, Anna K ; Lucek, Kay ; Maan, Martine E ; Marques, David A ; Martin, Simon H ; Matthews, Blake ; Meier, Joana I ; Möst, Markus ; Nachman, Michael W ; Nonaka, Etsuko ; Rennison, Diana J ; Schwarzer, Julia ; Watson, Eric T ; Westram, Anja M ; Widmer, Alex

Abstract: Speciation is a fundamental evolutionary process, the knowledge of which is crucial for understanding the origins of biodiversity. Genomic approaches are an increasingly important aspect of this research field. We review current understanding of genome-wide effects of accumulating reproductive isolation and of genomic properties that influence the process of speciation. Building on this work, we identify emergent trends and gaps in our understanding, propose new approaches to more fully integrate genomics into speciation research, translate speciation theory into hypotheses that are testable using genomic tools and provide an integrative definition of the field of speciation genomics.

DOI: https://doi.org/10.1038/nrg3644

# Genomics and the origin of species

Ole Seehausen[1,2#*], Roger K. Butlin[3,4#], Irene Keller[1,2,5#], Catherine E. Wagner[1,2#], Janette W. Boughman[1,6], Paul A. Hohenlohe[7], Catherine L. Peichel[8], Glenn-Peter Saetre[9]

Claudia Bank[10], Åke Brännström[11], Alan Brelsford[12], Chris S. Clarkson[13], Fabrice Eroukhmanoff[9], Jeffrey L. Feder[14], Martin C. Fischer[5], Andrew D. Foote[15, 28], Paolo Franchini[16], Chris D. Jiggins[17], Felicity C. Jones[18], Anna K. Lindholm[19], Kay Lucek[1,2], Martine E. Maan[20], David A. Marques[1,2,27], Simon H. Martin[17], Blake Matthews[21], Joana I. Meier[1,2,27], Markus Möst[17,21], Michael W. Nachman[22], Etsuko Nonaka[23], Diana J. Rennison[24], Julia Schwarzer[1,2,25], Eric T. Watson[26], Anja M. Westram[3], Alex Widmer[5]
(authors in this section ordered alphabetically)

*correspondence ole.seehausen@eawag.ch

[1] Department of Fish Ecology and Evolution, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Center for Ecology, Evolution and Biogeochemistry, Seestrasse 79, 6047 Kastanienbaum, Switzerland

[2] Division of Aquatic Ecology and Macroevolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland

[3] Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, UK

[4] Sven Lovén Centre – Tjärnö, University of Gothenburg, S-452 96 Strömstad, Sweden

[5] Institute of Integrative Biology, ETH Zürich, Universitätsstrasse 16, ETH Zentrum CHN, 8092 Zürich, Switzerland

[6] Department of Zoology; Ecology, Evolutionary Biology & Behavior Program; BEACON, Michigan State University, 203 Natural Sciences, East Lansing, Michigan 48824, USA

[7] Department of Biological Sciences, Institute of Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844-3051, USA

[8] Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA

[9] Department of Biosciences, Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, P. O. Box 1066, Blindern, N-0316 Oslo, Norway

[10] School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[11] Integrated Science Lab & Department of Mathematics and Mathematical Statistics, Umeå University, 90187 Umeå, Sweden

[12] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland

[13] Liverpool School of Tropical Medicine, Liverpool, UK

[14] Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, 46556-0369 USA

[15] Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, ØsterVolgade 5-7, DK-1350 Copenhagen, Denmark

[16] Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Universitätstrasse 10, 78457 Konstanz, Germany

[17] Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

[18] Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany

[19] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

[20] Behavioural Biology Group, Centre for Behaviour and Neurosciences, University of Groningen, P.O. Box 11103, 9700 CC Groningen, The Netherlands

[21] Department of Aquatic Ecology, Centre of Ecology, Evolution and Biogeochemistry, Eawag Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland

[22] Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Sciences Building, University of California, Berkeley, Berkeley, CA 94720-3160

[23] Integrated Science Lab & Department of Ecology and Environmental Science, Umeå University, 90187 Umeå, Sweden

[24] Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada

[25] Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

[26] Department of Biology, The University of Texas at Arlington, TX, USA

[27] Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland

[28] Current address: Dept of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

[#] core writing team

## Preface

Speciation is a fundamental evolutionary process, knowledge of which is critical for understanding the origins of biodiversity. Genomic approaches are an increasingly important aspect of this research field. We review current understanding of genome-wide effects of accumulating reproductive isolation and of genomic properties that influence the process of speciation. Building on this work, we identify emergent trends and gaps in our understanding, propose new approaches to more fully integrate genomics into speciation research, translate speciation theory into hypotheses that are testable with genomic tools, and provide an integrative definition of the field of speciation genomics.

## Introduction

Major insights into the genetics of speciation have come from a number of approaches (Box 1), ranging from the mapping of individual genes causing **reproductive isolation** (RI) to the characterization of genome-wide patterns of differentiation, and from quantitative genetic approaches to admixture analyses associating phenotypes with reduced **gene flow** between populations[1-3]. These empirical approaches have a long history, starting with the work of Dobzhansky[4] and Muller[5]. Theoretical understanding of the genetics of speciation has advanced markedly[6-10]. However, the deluge of empirical data coming from **next generation sequencing** (NGS), along with the emergence of new analytical approaches, necessitate the integration of this theoretical work to strengthen the conceptual foundations of the nascent field of **speciation genomics**. Such integration will help elucidate the relationships between evolutionary processes and genomic divergence patterns on the one hand, and between genomic properties and speciation processes on the other, and it will help unify research on the ecological and non-ecological causes of speciation.

In this review, we first discuss areas in which genomic approaches have begun to make important contributions to speciation research (Box 1), for example by elucidating patterns and rates of genome-wide divergence, improving our understanding of the genomic basis and evolution of intrinsic and extrinsic reproductive barriers, and identifying mechanisms by which different barriers become genomically coupled. We also highlight areas that would benefit from further attention; these areas include the distributions of **locus effect sizes, pleiotropy** and genomic constraint. We conclude by discussing how NGS data and innovative population genomic analyses could contribute to further progress in integrating these study areas into a more comprehensive and coherent understanding of the genomics of speciation.

## The evolution of reproductive barriers: Theory and classical evidence

In line with others[1, 3], we define speciation as the origin of reproductive barriers among populations that permit maintenance of genetic and phenotypic distinctiveness of these populations in geographical proximity. The origin of reproductive barriers can either be initiated by **divergent selection** (that is, "ecological" or sexual selection creating **extrinsic reproductive isolation**), or by the evolution - through genetic drift, as an indirect consequence of selection or through genomic conflict - of genetic incompatibilities that cause **intrinsic reproductive isolation** (Box 2). Studying the accumulation of intrinsic isolation has a strong tradition in evolutionary biology[1, 11]. Yet, most recent population genomic studies of divergence across the genomes of incipient and sister species have

investigated cases of putative **ecological speciation** and have focused on divergent adaptation and extrinsic isolation (but see[12] discussed below).

Extrinsic postzygotic isolation arises as a consequence of divergent or **disruptive** natural **selection** when the viability or fertility of migrants or of individuals with intermediate genotypes is reduced[2]. Prezygotic **sexual isolation** and also extrinsic postzygotic isolation, when hybrids have reduced mating success[13], may evolve as a consequence of divergent sexual selection[3, 14] which is often, but not always, mediated by differences in environments[15, 16]. Prezygotic sexual isolation and extrinsic **postzygotic isolation** are, hence, dependent on genotype-environment interactions in the wider sense (where mating partners are part of the external environment). In contrast, intrinsic postzygotic isolation is independent of the external environment. Consequently, different types of genes and gene networks and different evolutionary processes may be involved in generating these classes of isolation. Extrinsic postzygotic isolation and sexual isolation can evolve rapidly[17], and they often interact with each other[16] and with the evolution of intrinsic postzygotic isolating barriers[18] (Box 2). Selection can initiate speciation in situations with and without gene flow between populations, while intrinsic incompatibilities are less likely to accumulate when gene flow is present[6]. This being said, adaptive divergence and ecological speciation are not the same. Divergent adaptation alone rarely causes sufficient reproductive isolation to allow the accumulation or persistence of species differences in geographical proximity: this typically requires the evolution of **prezygotic isolation**[1, 3] (Box 2), although it is possible that this varies between major taxonomic groups such as insects versus vertebrates or plants.

The available evidence suggests that negative epistatic interactions, so called **Bateson-Dobzhansky-Muller incompatibilities** (BDMIs, or often just referred to as DMIs), are the most frequent cause of intrinsic postzygotic isolation[1, 19-21]. However, other mechanisms, including **underdominance**[22] and gene duplication, transposition and gene loss[23-25] may also cause intrinsic postzygotic isolation. The time course of the accumulation of DMIs is not well understood[19, 26-28], and rates may vary among taxa and among mechanisms underlying DMI evolution[19]. DMIs were long thought to arise either as a consequence of genetic drift, as a result of stochastic deactivation of gene duplicates[29] or as a by-product of ecological selection[30]. However, theoretical considerations, such as the slow pace of neutral accumulation of barriers[31], and early empirical evidence for positive selection on loci contributing to incompatibilities[32], suggested that drift was unlikely to be a common source of incompatibilities. Recent observations suggest instead that **intragenomic conflict** may be a common mechanism driving their evolution[20, 33-35] (Fig. 1), as originally proposed in 1991[34, 35]. Genomic conflict may arise from competing interests of males and females[36], from **meiotic drivers**[37, 38], mobile elements[39, 40], or other selfish genetic elements and their suppressors, and from competing interests between organellar and nuclear genomes[41, 42]. **Sexual conflict** is thought to drive the evolution of new sex chromosomes[43, 44], and empirical observations suggest sex chromosome turnover has a role in the evolution of reproductive isolation[45,46].

The different evolutionary mechanisms underlying the build-up of extrinsic and intrinsic postzygotic and of prezygotic isolation suggest that genomic signatures will also be distinct. The genomic architecture of extrinsic isolation is likely to resemble that of adaptive population divergence, and be diverse and scattered across multiple regions in the genome (see below). However, there are theoretical arguments and empirical evidence for spatial clustering of sites under selection in the genome when adaptive evolution proceeds under prolonged bouts of divergent selection with migration or recurrent hybridization[47]. For intrinsic isolation, incompatibility factors driven by genomic conflict are expected to accumulate in genomic regions of reduced recombination where linkage disequilibria between **distorter loci** and **responder loci** can become established[48, 49]. Sex chromosomes are particularly susceptible to the accumulation of incompatibility factors derived from genomic conflict because sex chromosomes are constantly in a battle over segregation, whereas only small and tightly linked autosomal regions are in conflict with their homologs[34]. At the same time,

there will be particularly strong selection for suppression of sex-linked distorter loci because they tend to bias sex ratios[50, 51]. The genomic architecture of certain types of prezygotic isolation may also be influenced by regions of reduced recombination around sex determining loci[52] or sex chromosomes[53], particularly when sex-linkage resolves sexually antagonistic effects of sexual selection[54]. Alternatively, prezygotic isolation loci may accumulate near extrinsic ecological isolation loci (see section below, "***Genomic coupling of reproductive barriers"***). All of these signatures must be distinguished from background patterns of genetic diversity and divergence that depend on the populations' history of genetic drift, gene flow, background selection and episodes of positive selection unrelated to reproductive isolation.

Looking for signatures in the genetic architecture of reproductive isolation has a long "pre-genomic" history ([55, 56]). However, there has been a historical disconnect between research programs focused on intrinsic isolation, which have typically concentrated on later stages of speciation[20, 57], versus extrinsic postzygotic isolation and prezygotic sexual isolation at early stages of speciation[2, 30,15, 16]. Because of this disconnect, comparing the rates of evolution of components of reproductive isolation, and their relevance to speciation, is currently a challenge. Where rates have been compared in the same taxon using "pre-genomic" methods[11, 58-60], the data suggest that prezygotic and extrinsic postzygotic isolation often evolve faster than intrinsic postzygotic isolation, consistent with expectations from classical theory[61]. Genome-wide data will now permit testing of this pattern with a tremendous increase in resolution.

## Genomics and the speciation continuum

Once speciation is complete, populations accumulate differences due to mutation and genetic drift as well as ongoing selection. Reproductively isolated species, therefore, often differ in traits that evolved under ecological selection and others that evolved under sexual selection, and may also have intrinsic incompatibilities. A central task of speciation genetics is to reconstruct the sequence in which these different barriers originated so as to distinguish between causes and consequences of speciation. To achieve this, one would ideally take an unbiased view of the entire genome at all stages of the same speciation process. However, speciation can rarely be studied in real time in natural populations of sexually reproducing multicellular organisms. Estimates of variation among loci in the timing and magnitude of gene flow could help determine the order in which reproductive barriers emerged, but such inferences are challenging and current methods are not accurate enough for this purpose[62]. However, by integrating case studies of closely related taxa that vary in their extent of divergence (the "speciation continuum"), inferences can often be made about the chronology and significance of different factors and processes at play.

Investigations of this **"speciation continuum"** have made important contributions to speciation research[63, 64] and this approach is being adopted in NGS-based genome and transcriptome scan studies of speciation. The major questions being addressed are the extent to which divergence at different stages in speciation is either localized in the genome (the "island view") or widespread, the extent to which heterogeneity in divergence can be attributed to selective processes versus genetic drift, the sources of selection, whether genomic divergence tends to follow a common trajectory as it proceeds along the speciation continuum, and how all this is affected by the extent of geographical isolation. A recently much cited scenario for speciation without strong geographical isolation, derived from earlier models[65, 66], involves an early stage of divergence where differentiation is limited to a small number of loci (islands) under strong divergent selection. Gradually, these regions would grow through the process of **divergence hitchhiking**, and eventually the effective migration rate would become reduced globally across the genome fostering genome-wide divergence ('genome hitchhiking')[67, 68].

***Genome scans of ecological speciation***

Several NGS-based **genome scans** of the speciation continuum have found surprisingly variable patterns of genomic divergence. It appears that incipient species can quickly accumulate substantial divergence, even in the presence of gene flow (Fig. 2). However, whereas in some examples - such as *Heliconius* butterflies[69], *Helianthus* sunflowers[70], and poplar trees[71] - divergence between **parapatric** ecotype populations is limited to a few large genomic regions, in others it is widespread across the genome[72-75]. NGS-based genome scans of **sympatric** sister species have generally reported genomically widespread and highly heterogeneous divergence that varies on a very local scale[75-81]. Few studies have looked for evidence of divergence hitchhiking and the available results are inconsistent[69, 76, 82]. Genome-wide average $\mathbf{F_{ST}}$ often increases as phenotypic divergence increases[80, 83] but divergence seems to remain heterogeneous across the genome for a very long time, potentially due to repeated episodes of interspecific gene flow even after RI has become strong[84, 85]. The first generation of NGS-based population-genomic studies of ecological speciation has therefore shown that ecological selection can cause strong isolation of small genomic regions between diverging populations, and that when RI is strong enough to permit persistence of incipient species in sympatry, many unlinked regions typically experience significant isolation.

So where does the heterogeneity in genomic divergence come from? It is commonly inferred to result from locus-specific differences in the effects of divergent selection and gene flow. Indeed, genome scans have shown strong isolation at genomic loci that were known to be under divergent selection[64, 69, 70, 72, 74]. However, caution is warranted as different evolutionary processes can leave similar signatures in the genome. Heterogeneous genomic divergence is sometimes also observed between allopatric populations of the same species in the absence of any current gene flow[76, 86, 87] (Fig. 2). Indeed, many studies assume ongoing gene flow between species, even though stochastic variation due to recent **coalescence** times and **incomplete lineage sorting** can similarly lead to low divergence and high heterogeneity, particularly when in combination with selection[88, 89]. Statistical methods are available to distinguish divergence in isolation from divergence with gene flow, and these methods are increasingly being applied to genome scale datasets (reviewed in [90]; Box 1).

Even in the absence of selection, divergence is expected to vary due to the stochasticity of genetic drift and the complexities of population history, and this variation can be enhanced by confounding effects of genomic heterogeneity[91]. In particular, regions of low recombination and/or high gene density often show reduced intra-specific diversity, which inflates relative divergence as measured by $F_{ST}$ or $D_a$[88]. This can result from background selection against deleterious mutations[92], intraspecific selective sweeps (in allopatry)[88] or even a direct influence of recombination on genetic diversity[93]. Disentangling these processes is challenging[94]. Some have suggested correcting for recombination rate in interpreting $F_{ST}$ patterns[83]. Others have suggested that absolute divergence measures such as $\mathbf{D_{xy}}$ are more robust to diversity artefacts[95], especially when corrected for local mutation rate[96]. It seems unlikely that any single parameter will reliably disentangle divergent selection and gene flow from neutral processes. Good knowledge of the geographical context of population divergence will help, but distinguishing between hypotheses of speciation with gene flow, secondary contact and incomplete lineage sorting will often require new, parameter-rich modeling approaches[90].

Adaptive divergence has been shown to accumulate preferentially in regions of low recombination[97], including the centers of chromosomes[83], the vicinity of centromeres[98], inversions[74] or often (but not always[12, 71]) on sex chromosomes[98-100]. Heterogeneity in genomic divergence seen in allopatry might also result from **gene-flow-selection balance** that has occurred in the past[47, 76]. Finally, the assumption that the baseline $F_{ST}$ reflects neutral divergence may be violated in cases where divergent selection is pervasive and multifarious, and this would bias against the detection of the signature of selection[81].

Evidence for repeated divergence of the same genes or genomic regions across replicate pairs of species or environmental contrasts already provides strong evidence that these regions are indeed involved in adaptation and/or RI[72, 74, 85, 97, 101-103]. Detecting such parallel divergence may require dense sampling of genomes or transcriptomes because the highest levels of repeatability may be observed at the scale of genomic regions rather than individual genes or SNPs[97]. In this case, the repeatability in the heterogeneity of genomic divergence may be due at least in part to shared genomic heterogeneity in recombination and mutation rates rather than parallel adaptive divergence, but the shared genomic structure may facilitate the repeated accumulation in the same genomic regions of adaptive differentiation[97]. Another approach involves combining classic **cline** theory with genome-wide analyses, allowing measurements of the strength of selection at specific loci[79] (Box 1). In the future, parameter-rich **coalescent** models of divergence with gene flow fitted to genomic data may be able to account for the heterogeneity of demographic history across the genome when seeking to identify genomic regions with reduced gene flow[104, 105]. Finally, genome scans combined with manipulative selection[81], QTL mapping[82, 106], candidate gene mapping[72, 74] and **admixture mapping**[79, 107-109] can be used to investigate whether divergent genomic regions contain loci contributing to RI.

Several recent studies have found a contribution of ancient alleles to recent divergence, as exemplified by stickleback[74, 110], cichlids[77, 111], *Rhagoletis* flies[112] and *Heliconius* butterflies[113]. Ancient alleles are identifiable due to the accumulation of many substitutions or sharing over wide spatial or taxonomic ranges. The sources of such ancient allelic variation can either be **standing genetic variation**, or **hybridization**[114]. Distinguishing between these hypotheses is difficult in practice due to the challenges of distinguishing **incomplete lineage sorting** from hybridization[90] (Box 1). The balance of evidence from NGS data implies **introgressive hybridization** rather than standing variation as the source of ancient alleles in most of the above cases. Speciation in these cases might have been facilitated by hybridization providing genetic material for adaptation and reproductive isolation in the face of gene flow, which puts a new twist on an old idea [1]. Future research combining genomic and ecological approaches should test this hypothesis further.

***Genomic divergence and intrinsic isolation***

Many studies have investigated DMI genes in strongly isolated species, but in many cases it remained unclear if the **fixation** of the underlying mutations was a cause or a consequence of speciation[20, 57]. Regardless of whether identified DMI alleles are the first step in the origin of reproductive isolation, a striking pattern to emerge from recent work is that they have evolved under strong positive selection rather than genetic drift and that **genomic conflict** is often implicated as the source of this selection. For example, one study identified *Ovd*, an X-linked gene that underlies both hybrid male sterility and sex-ratio distortion in crosses between *Drosophila pseudoobscura pseudoobscura* and *D. p. bogotana*[51]. Another example is a recent analysis that found strong evidence for ongoing positive selection within *Drosophila mauritiana* in genes that have diverged between this species and its closest relatives and that are known to be involved in genomic conflict[12]. Two pronounced polymorphism troughs on the X chromosome were centered on a pair of genes that cause sex-ratio distortion within *D. simulans*, and on *Odysseus*, a rapidly evolving homeobox gene that was known to cause male sterility in *D. mauritiana* x *D. simulans* hybrids[32] and may be involved in genomic conflict. These are two candidate cases of speciation by conflict-driven DMI evolution.

***Genomic coupling of reproductive barriers***

The build-up of associations between several traits or loci involved in RI strengthens the total barrier to gene flow between diverging populations, and is therefore important for the evolution of strong reproductive isolation[115, 116]. Such **genomic coupling** can involve any pre- or post-zygotic barriers[117]. Deviations from linkage equilibrium between barrier loci can initially be generated by new mutations arising on a particular genetic background, or by genetic drift during divergence with limited gene flow. Coinciding barriers may, for example, arise through secondary contact between divergent

populations, through the evolution of DMIs as an incidental by-product of divergent selection[118], or via hitchhiking of intrinsic incompatibility alleles with divergently selected alleles, as has been shown for heavy-metal adapted populations of monkey flowers[119]. However, for barrier coupling to be important in speciation, coupling has to be maintained or even strengthened in the face of gene flow, and this typically requires divergent selection[6].

Selection is expected to favour the coupling of barriers if this leads to an increase in mean fitness. In theory this can involve multiple intrinsic barriers (like DMIs)[120, 121] or intrinsic and extrinsic postzygotic barriers as well as sexual and other prezygotic isolation traits. Across an **ecotone**, multifarious extrinsic selection can assemble and maintain many coinciding clines at loci involved in adaptation[122], and these can become coupled with sexual isolation traits[123] and with DMIs[18,116, 124]. Selection can also directly favour the evolution of increased prezygotic isolation, as in the case of **reinforcement**[125]. Finally, sexual conflict can couple intrinsic postzygotic and prezygotic sexual isolation because DMIs driven by sexual conflict and genes underlying sexual traits or preferences expressed only in one sex may both accumulate on sex chromosomes[53, 126]. Consistent with these expectations, loci for plumage colour, mating preferences and intrinsic postzygotic incompatibilities are coupled on the *Z* chromosome in flycatchers[52] and Gouldian finches[127, 128]. Similarly, loci for behavioural isolation and hybrid male sterility are coupled on the *X* chromosome in a species pair of Japanese stickleback[45].

Because recombination tends to break up gene associations, genomic architectures that eliminate or decrease recombination are expected to facilitate coupling, and hence speciation[129]. Most prominently, recombination will affect neither associations among traits that are pleiotropically influenced by the same allele, nor **'one-allele' mechanisms**, where the presence of the same allele in different genetic backgrounds confers RI[130]. One-allele mechanisms do not leave a population-specific signature in the genome at the primary isolation locus but they should be detectable as **sweeps** shared by both diverging populations if they arise during speciation (as for instance if an allele for imprinting on the phenotype of the father spreads across two incipient species that were connected by gene flow). Despite the theoretical expectation that 'one-allele' mechanisms evolve more readily during speciation with gene flow than other types of barriers[6, 16, 130], we are not aware that the predicted genomic signature of shared sweeps at isolation loci has yet been detected in any case. Revealing such a signature would be a strong contribution of speciation genomics to demonstrating a classical prediction of speciation theory.

Loci underlying **'two-allele' mechanisms** are expected to be concentrated in regions of reduced recombination. Recent genomic studies have observed genomic architectures that eliminate or reduce recombination between traits involved in RI: There is evidence of synergistic pleiotropy in **multiple-effect or "magic" traits**[16, 131-133], and multiple genes underlying isolating traits have been found together in inversions[134-136], on sex-chromosomes[45, 52, 128] and also in otherwise tight physical linkage[119, 137], including mating traits and mating preferences in cases of speciation with gene flow[138]. These data also provide some evidence that reinforcement of prezygotic isolation is facilitated by linkage, as in flycatchers[139], or by pleiotropy, as in phlox[132]. In other cases reinforcement might be constrained[140] where loci are not linked and where there is extensive gene flow. However, recent genomic studies have also provided empirical examples of coupling between unlinked loci in fully **sympatric** hybridizing species[77] and especially in **hybrid zones**, where clines at many unlinked loci often coincide, although it is not always clear exactly how these loci are implicated in RI[141]. Unbiased whole-genome re-sequencing data and genome scans from diverging populations, coupled with methods to reduce bias from NGS data[142] and with mapping of isolation traits, are needed to test the generality of these patterns.

***Effect sizes and pleiotropy***

A key question, with a long history[55, 143], is whether speciation is typically initiated by divergence at few loci of large and possibly pleiotropic effect or by divergence at many loci with small and additive

effects[133, 144]. The distinction is important because it will affect how speciation is constrained by the availability of suitable genetic variation, and will also affect how likely it is that selection or genetic drift may overcome gene flow. On their own, $F_{ST}$ estimates from genome scans tell us little about the effect sizes of individual alleles on phenotypes, fitness or RI[107] (Fig 3). With regard to fitness, Fisher's geometric model predicts that the probability that a mutation is favourable decreases exponentially with mutational effect size, hence we expect few alleles of large positive fitness effect but many of small effect[145-147] (but see[148]). However, this prediction does not take into account standing genetic variation, gene flow or changing environments. When those factors are considered, the predictions change[47, 147, 149] and may even reverse[150].

Speciation with gene flow may require divergent or disruptive selection to be concentrated on a small number of regions in the genome that also have large effects on RI[6]. Theoretically expected distributions of effect sizes in terms of RI (rather than fitness) may be different for different classes of isolating barriers, but current data are equivocal (Fig. 3b). For example, mapping hybrid inferiority in natural environments for *Arabidopsis* has shown RI to be due to many genes with moderate effects[151]. In contrast, hybrid inviability in *Mimulus guttatus* is a consequence of two linked loci of major effect[119]. Predictions about the distribution of effect sizes expected for genes that underlie DMIs are also generally lacking, partly because effect sizes depend on mutation order and the extent of background genomic divergence. Traits governing prezygotic isolation, and especially sexual isolation (Box 2), are likely to have large effects on RI because they directly influence mating or fertilization patterns[1, 6, 16, 152-154]. To test this prediction with genomic scale data, existing quantitative genetic, mapping and candidate gene studies[45, 109, 111, 128, 138, 155-157,158, 159] should now be followed up by NGS-based genome scans assessing RI around these loci[107].

Recently identified large-effect alleles involved in adaptation and speciation with gene flow, are often highly **pleiotropic** (e.g., *Optix* in *Heliconius*[160] and Ectodysplasin [*Eda*] in sticklebacks[161], although we lack estimates of the effect *Eda* has on RI or fitness). Such alleles may be rare among newly arising mutations but alleles with synergistically pleiotropic effects may be more common in standing genetic variation. Recent theory suggests that large-effect or pleiotropic alleles may be favoured by selection during evolution in gene-flow-selection balance, and hence eventually become enriched in taxa with divergence and gene flow[47].

***Genomic constraint***

The flipside of the coupling problem is that genetic correlation between traits as a result of pleiotropy or tight linkage may also constrain speciation. With new population genomic data revealing divergence in many regions of the genome early in speciation, there is an opportunity to unite population genomics with a quantitative genetics perspective on the evolution of polygenic traits during speciation. In quantitative genetics terms, standing genetic variation is quantified by the **G-matrix** of additive genetic variance and covariance[162]. **G** may indicate potential constraints on adaptive evolution that affect the response to directional selection[163, 164], as well as constraints on genetic drift[165]. Tests to detect the impact of selection on G are available (e.g.[166]). Divergence among populations is biased along axes with greater genetic variation and covariation and constrained along axes with little variation or covariation[164, 167, 168]. Importantly, however, genetic constraints are not only negative. Genetic covariation may align with **correlational selection**[169, 170] and, as discussed above, pleiotropy can couple adaptation to RI. It is not known how genes of major effect, versus the traditional assumption of many genes of small effect, influence the structure of **G**[171], and how higher moments of the distribution of genetic variation and covariation affect the response to divergent selection[172]. These questions can now be addressed with genomic methods, such as directly estimating **G** in outbred populations using NGS data[173]. A different approach is to estimate the genetic variance–covariance matrices for gene regulatory networks from gene expression data. Analyzing genomic data in a quantitative genetics framework in this way will illuminate how genomic constraint affects speciation[174].

Studying effects of hybridization is one promising application. Beyond being a source of allelic variation, hybridization may facilitate evolution and perhaps speciation by releasing populations from constraints caused by genetic correlations. While empirical evidence has accumulated that suggests that selection alters genomic architecture[169, 175], the role of gene flow in aligning **G** with the direction of divergent or disruptive selection has rarely been investigated[150]. The emerging consensus that hybridization frequently introduces adaptive variation[18] calls for empirical studies in this area. We predict that hybridization will influence speciation not only by generating novel and **transgressive phenotypes** but also by aligning **G** with the axis of divergent selection (Fig. 4a). Even when early generation hybrids are maladapted, hybrid populations may over time benefit from increased evolvability[176]. Hybridization may alter patterns of genetic covariance much faster than is possible by selection alone, and may lead to bursts of evolutionary diversification and speciation[114, 177] (Fig. 4b-d). Genomic methods can now be used in assessing these hypotheses in several ways, such as direct estimation of **G** in both parental and hybrid natural populations and through association or **admixture mapping** of loci contributing to novel adaptive phenotypes in hybrid populations[108].

## Speciation genomics: towards a synthesis

Speciation can proceed in many different ways, but these can be grouped in terms of drivers (drift and different types of selection), causes (extrinsic environment-dependent versus intrinsic environment-independent) and stage in the life cycle (postzygotic or prezygotic) of reproductive isolation, resulting in two major classes that are at least in theory quite distinct (Box 2). In one, RI is initiated by extrinsic selection, in the other by intrinsic incompatibility. Analysis of NGS data has begun to shed light on the signatures of these processes in the genome. Both of these classes of processes can generate reproductively isolated species in allopatry, but parapatric and especially sympatric speciation are constrained to situations where divergent natural and/or sexual selection overcome the homogenizing effects of gene flow[1, 6]. Whether speciation in such scenarios can proceed depends on the strength of selection[2, 6] and the genetic architecture of adaptation and reproductive isolation[76, 122]. Speciation driven by genomic conflict is much less likely to be initiated in the presence of gene flow because selfish genetic elements may then spread across populations and thereby prevent or slow down the accumulation of conflict-driven DMIs[178]. However, it remains possible that relatively brief periods of **allopatry** are sufficient for the origins of conflict-driven DMIs. Although DMIs may be removed by selection after **secondary contact**, they may, in theory, facilitate speciation if they become coupled with other components of RI before they are purged[116, 179]. How often this happens is unknown.

These principles are not new[1], but they can and should now be examined with much greater resolution using genomic methods. Although speciation genomics is clearly still in its infancy, a few trends are emerging from the first generation of NGS-based genome scans, particularly in relation to non-allopatric speciation: The available evidence suggests that divergence can be genomically widespread very early in speciation, and may generally be so in species that coexist in full sympatry[74-77, 80], whereas it can be restricted to very few islands of divergence in parapatric ecotypes[69, 70]. Perhaps **multifarious divergent selection** or genomically widespread selection is important to generate sufficient RI to permit maintenance and perhaps buildup of genetic differentiation in sympatry. More data are now needed to confirm this intriguing pattern.

Some genomic regions that are divergent between incipient and sibling species in geographical proximity contain genes with large effects on adaptation and pleiotropic effects on prezygotic isolation. The alleles at several such loci have turned out to be ancient variants that were present as standing variation or were brought together by hybridization in the ancestors of emerging species

pairs[99, 111, 112]. Although it is premature to draw strong conclusions, this may turn out to be another emergent feature of speciation with gene flow. We expect effect sizes to be larger, antagonistic pleiotropy to be less frequent and synergistic pleiotropy to be more frequent in ancient alleles that have been honed by selection over time than in alleles arising newly through mutation. We hypothesize that substitution of such ancient alleles at major effect loci has the potential to reduce gene flow quickly, to the point where substitutions with smaller effects at other loci can also spread. Genome scans of divergence very early in the speciation continuum (at low overall RI, Box 2) should allow explicit tests of these hypotheses.

Alternative mechanisms and geographical modes of speciation make different predictions for patterns in genomic data. Specifically, we predict that speciation due to conflict-driven DMIs involves greater divergence at centromeres and sex chromosomes, and so these regions should bear signatures of selective sweeps. Divergence under ecological selection may be more widely distributed across the genome, and sweeps at individual loci less pronounced. The available data are consistent with these expectations, although theory predicts accumulation of genes for ecological divergence in regions of low recombination when selection is antagonized by gene flow[129]. Divergence by sexual selection may be concentrated on sex chromosomes[52], but support for this prediction is not always found and predictions vary with the sex determination system[54]. Many more population genomic studies of divergence in a wider range of taxa and across a greater range of points along the speciation continuum are needed to test these predictions further. Speaking more broadly, future work should seek to determine to what extent different evolutionary mechanisms and geographical modes of speciation can be distinguished based on genomic data and, in turn, the extent to which genomic features can predict the modes and mechanisms of speciation that apply to a given evolutionary lineage.

Taxonomic variation in the propensity for speciation without geographical isolation is prevalent[180] and it will be interesting to learn if variation in genomic architecture explains some of this. Whether selection can overcome gene flow depends, besides the total strength of selection, on the number of genome regions targeted by selection, on the rate of recombination between them, and on the extent of pleiotropy. When analyzed in conjunction with ecological data, genomic data therefore hold promise to help explain why non-allopatric speciation occurs readily in cichlid fish, whitefish, stickleback, *Rhagoletis* flies, *Heliconius* butterflies, *Coprosma* shrubs[181] and some other taxa, but is not reported in the majority of others. This combination of approaches may also help more generally to explain why some taxa undergo speciation and accumulate species diversity a lot more readily than others. Answering such questions will also facilitate an understanding of larger-scale patterns in species diversity (Box 3).

Population-genomic studies that explicitly compare rates of evolution and the genomic distribution of prezygotic, extrinsic postzygotic and intrinsic postzygotic barriers to gene flow have yet to materialize. We believe that such studies hold considerable promise to overcome old dichotomies in speciation genetics. Because the discovery of DMIs used to be laborious, we cannot yet answer the question how often DMIs are caused by conflict, ecological selection or genetic drift. This too will hopefully soon change as genomic data allow the identification of DMI loci at an increasing pace[12, 26] (Box 1).

A still missing part of a synthesis in speciation genomics is the integration of evolutionary developmental biology. Insights from this field make several relevant suggestions for speciation genomics[182]: First, mutations in coding sequences may more often have pleiotropic effects than those in *cis*-regulatory regions. Second, pleiotropy will be more common when selection targets genes with central roles in gene regulatory networks, and many morphological developmental genes are in such positions. Third, because of the first two predictions, morphological evolution may often be constrained to take place through changes in *cis*-regulatory mutations, whereas physiology may

be more free to evolve through coding mutations. These predictions make for interesting yet little explored connections between some of the above discussed questions in speciation research and the debate about the prevalence of coding versus *cis*-regulatory mutations in evolution[182, 183]. Possible ascertainment bias notwithstanding, empirical data suggest that divergence between sibling species and conspecific populations is predominantly due to evolution of coding genes, independent of their positions in gene regulatory networks, but morphological differences between species that diverged longer ago are predominantly due to *cis*-regulatory evolution[182]. The following explanation has been offered: Selection acting early during population divergence may partly overcome the negative fitness effects of antagonistic pleiotropy that are expected for coding mutations, but may not be strong enough to fix these mutations[182]. Over time, as more mutations become available, *cis*-regulatory mutations with more specific effects and less antagonistic pleiotropy would replace the coding variants. An interesting implication is that the mutations responsible for phenotypic differences between older species may be distinct from those that are causally important in the process of population divergence and speciation, even when the mechanism of speciation and the diverging phenotypes are the same. Studies of the genomic basis (coding versus regulatory) of species divergence in incipient versus older species in the same taxon are needed to test this hypothesis. We are not aware that such data exist.

These are exciting times for speciation research, and major progress in the field is likely to come from integrating the analyses of genomic data with studies of ecology, behavior, developmental biology and theory. We propose three major building blocks as a roadmap for such continued integration.

First, there is a need for more comparative genome scans at different stages in the speciation continuum in closely related taxa or in replicate species pairs in the same taxon. These data need to be combined with annotation of the effects of alleles on phenotypes and on RI, which can be done through QTL mapping or functional analyses in the context of annotated reference genomes. This would allow the association of divergent genomic regions with mechanisms of RI. Such studies need to be repeated in the following scenarios: in taxa in which speciation is driven by ecology, sexual selection and intrinsic incompatibilities (Box 2); in different spatial contexts; and in taxa that have not speciated, but that occupy similar environments to those taxa that have undergone speciation. Sampling design should explicitly aim to explore variation, both in different stages on the speciation continuum and for different degrees of geographical isolation (Fig. 2), and the history of geographical isolation should ideally be known. Eventually, with replication and clever experimental and comparative study designs, it will become possible to understand whether different mechanisms and modes of speciation can be distinguished based on patterns observed in genome-wide data.

Second, experimental **population genomics** studies of speciation are needed to measure the strength and multifarious nature of selection, and more generally to test hypotheses about processes underlying differentiation and isolation, including intragenomic conflict, heterogeneity in recombination rates, and coupling.

Third, theoretical modeling is needed that includes the influences of variable demography, recombination rates and time, and explicitly considers standing genetic variation and different sources of incompatibilities. Such models will be helpful in generating predictions that can be tailored to individual empirical study systems to make them testable. Such predictions could include genomic signatures of alternative speciation modes and mechanisms, and how modes and mechanisms can be inferred from patterns found in genomes at different stages of the speciation continuum. Improved methods for estimating the timing of long-term gene flow would also be very valuable[90]. Given the increasingly widespread evidence for recruitment of ancient genetic variation into recent speciation events, analytical methods for rigorous hypothesis-testing regarding the source of such variation – that is, the contributions of

hybridization and standing genetic variation – are also needed. Such methods could include comparisons of the phylogenetic histories of genomic regions that confer adaptation and reproductive isolation with those of other segments of the genomes of young sister species[74, 77, 99, 112].

## Conclusions

New approaches for gathering large amounts of genomic data in non-model organisms have begun to produce intriguing and unexpected insights into the genetics of speciation. Sympatrically coexisting species are characterized by heterogeneous differentiation that is widely scattered across the genome even when these species are still very young, but adaptive differentiation between parapatric populations can be restricted to a few genomic islands. Ancient alleles with large and pleiotropic effects characterize both types of divergence, and were often acquired by interspecific hybridization. Genomic conflict may be a frequent source of intrinsic postzygotic isolation. It may be recognized in genome scans as strong sweep signatures on sex chromosomes or in isolated islands of divergence on autosomes. More strongly integrated studies are now needed that cover multiple components of RI at multiple stages of the speciation continuum, and in geographical settings ranging from complete allopatry to full sympatry, paying additional attention to the history of population contact (primary or secondary). With the rapid growth of genomic data generation and analysis approaches, it will then soon become possible to construct an integrated picture of speciation starting from the evolution of reproductive barriers and how this is influenced by ecological and genomic constraints, through the way speciation creates signatures of genomic divergence, to how genomic properties of organisms interact with history and ecology in shaping patterns in biodiversity. There is no doubt that a new phase of discovery has begun that will usher in a greatly increased understanding of the origin of species.

Author contributions. This paper was initiated during a workshop funded by the ESF networking program Frontiers in Speciation Research (FroSpects). OS led workshop organization and coordination, and manuscript preparation with assistance from CEW, IK and RKB. JWB, PAH, CLP, G-PS, CEW and IK led discussion groups and initial drafting of sections of the paper, ETW, CDJ, CSC, SHM, JWB, JS and CEW prepared figures, OS and RKB drafted general sections, other authors contributed during the workshop and commented on drafts.

Acknowledgements. We thank the European Science Foundation (networking program Frontiers in Speciation Research, FroSpects) for funding a workshop on “Genetics and Genomics of Speciation” and for contributing towards publication costs, and Nadja Pepe and Lucie Greuter for help with the organization of the workshop. We thank Christian Lexer and two anonymous reviewers for constructive suggestions that improved our paper.

**Glossary**

Items included in the glossary are bolded in their first appearance in the main text.

**Admixture mapping**

Identification of genetic loci that contribute to phenotypic differences between ancestral populations, by exploring genotype-phenotype correlations in a population of mixed ancestry.

**Allopatric**

Organisms, populations or species inhabiting distinct geographical regions and therefore not exchanging genes.

**Allopatry**

Occurrence in geographically isolated regions.

**Cline**

Directional variation in phenotype or genotype, or change in frequency (e.g. of an allele), across a geographic region.

**Coalescence**

The merging of two genetic lineages in a common ancestor.

**Coalescent**

A statistical framework for the analysis of genetic data where the genotypes shared by populations or species are traced back in time to their most recent common ancestor.

**Correlational selection**

Selection for optimal character combinations.

**Disruptive selection**

Selection within a single population that favours extreme phenotypes over intermediate phenotypes.

**Distorter loci**

Loci underlying **meiotic drive**, the non-Mendelian segregation of alleles in meiosis. Distorter loci may act on other loci, so-called responder loci.

**Divergence hitchhiking (DH)**

Occurs when divergent selection on a locus reduces the effective migration rate for physically linked regions, which increases the opportunity for divergence at loci under weaker selection in these surrounding regions. DH regions may remain much larger than traditional hitchhiking regions after a selective sweep within populations because of the persistent reduction in the ability of flanking regions to recombine away from a divergently selected gene.

**Divergent selection**

Selection favouring different phenotypes in different populations.

**$D_{xy}$**

The average number of nucleotide substitutions per site between two populations.

**Bateson-Dobzhansky-Muller Incompatibility (BDMI or mostly just referred to as DMI)**

An intrinsic postmating barrier that is the result of epistatic interactions between alleles at two or more loci that cause reduced fitness in hybrids but not in the parental populations.

**Ecological speciation**
The evolution of reproductive isolation as a consequence of divergent or disruptive natural selection between populations that inhabit different environments or exploit different resources.

**Ecotone**
A zone where there is a transition between two distinct biological communities, e.g. between forest and grassland or aquatic and terrestrial habitats. Ecotones are typically associated with changes in the physical environment.

**Extrinsic reproductive isolation**
Fitness reduction in hybrids that is dependent on the environment, i.e. mediated by genotype-environment interactions.

**Fixation**
Describes the situation in which a mutation or variant has achieved a frequency of 100% in a population.

**$F_{ST}$**
A measure of population subdivision that compares the correlation between two gene copies that are randomly drawn from the same population to that between two gene copies drawn from two different populations. An $F_{ST}$ of 1 indicates that two populations are fixed (**fixation**) for alternative alleles.

**$F_{ST}$-outlier analysis**
Comparison of the distribution of $F_{ST}$ values across loci with the distribution expected in the absence of divergent selection for the same average differentiation. Loci whose $F_{ST}$ values exceed expectation are likely to be influenced by divergent selection, either on the locus itself or on a linked locus.

**Gene flow**
The movement of alleles between populations. For gene flow to occur, individuals must disperse between populations and successfully reproduce with local individuals. Therefore, gene flow can be reduced not only by dispersal barriers but also by intrinsic or extrinsic reproductive isolation.

**Gene-flow-selection balance**
A level of differentiation between sub-populations at which the homogenizing effect of gene flow and the differentiating effect of divergent selection are in equilibrium.

**Genome scan**
Comparison of genome-wide patterns of diversity within populations and/or divergence between populations at hundreds or thousands of markers. Most studies until recently used Amplified Fragment Length Polymorphisms (AFLPs) but this has recently changed, and SNPs generated by NGS or SNP chips are being used.

**Genomic conflict**
Genomic conflict arises between genes or genetic elements within the same genome when these are not transmitted by the same rules (e.g. biparental vs uniparental inheritance), or when a gene causes its own transmission to the detriment of the rest of the genome. The presence of elements (distorter loci) that bias transmission is expected to lead to the evolution of loci that restore Mendelian segregation (restorer loci).

**Genomic coupling**
The statistical association between different traits and loci involved in RI.

**G-matrix**
The additive genetic variance-covariance matrix that summarizes the variances within and covariances between multiple phenotypic traits.

**Hybridization**
Mating between individuals that belong to distinct species or populations. If postmating isolation is incomplete, hybridization leads to the introgression of genes from one population to the other.

**Hybrid zones**
Spatially restricted regions where the distribution ranges of distinct populations or incipient species come into contact and hybrids are formed.

**Incomplete lineage sorting**
Situation in which some alleles share a more recent common ancestor with alleles in another species than with other alleles in the same species.

**Intragenomic conflict**
Antagonistic selection among genomic elements with different fitness interests in an individual.

**Intrinsic reproductive isolation**
Fitness reduction in hybrids that is independent of the environment.

**Introgressive hybridization**
The introduction of genes from one population or species into another through hybridization.

**Linkage disequilibrium**
The statistical association of the alleles at two loci within gametes in a population. Although linkage disequilibrium tends to be greater between linked loci, it can also arise between physically unlinked loci — for example, because of selection, non-random mating or gene flow.

**Locus or allele effect size**
The magnitude of the influence of a locus, or a specific allele, on a phenotypic trait. This can be expressed, for example, as the proportion of phenotypic variation attributable to a specific locus or the phenotypic difference between genotypes with and without a specific allele.

**Multifarious divergent selection**
Divergent selection acting on multiple traits.

**Multiple-effect traits or “magic” traits**
Traits that contribute to more than one component of reproductive isolation, such as a trait contributing to local adaptation that is also used as a mating cue.

**Meiotic drivers**
Factors distorting Mendelian segregation. At a heterozygous site, the driving variant will be found in more than half of the gametes.

**Next Generation Sequencing**
A class of high-throughput sequencing methods that rely on technologies that parallelize the sequencing process, producing thousands or millions of sequences concurrently. Next Generation Sequencing technologies increase throughput and lower the cost of DNA sequencing by orders of magnitude compared to standard dye-terminator methods.

**One-allele mechanism**
Reproductive barriers arise through spreading of the same allele in each of two diverging populations, such as an allele for behavioural imprinting or reduced migration.

**Parapatric**
Organisms, populations or species that inhabit adjacent geographical regions or spatially distinct but adjacent habitats and may exchange genes.

**Pleiotropy**
Effect of an allele on more than one trait.

**Prezygotic isolation**
Effect of barriers acting before or after mating but before fertilisation, including the isolating effects of divergent mate choice, habitat preference, reproductive timing and gametic incompatibility.

**Population genomics**
Use of genome-wide data (typically based on next-generation sequencing methods) to make inferences about evolutionary processes in natural populations.

**Postzygotic isolation**
Effects of barriers acting after fertilisation, such as hybrid sterility and hybrid inviability. Can be extrinsic (mediated by the environment) or intrinsic.

**Quantitative trait locus (QTL)**
Chromosomal region with a statistically significant effect on a phenotype.

**Reinforcement**
Selection for the strengthening of prezygotic barriers to avoid the production of unfit hybrids between taxa that have previously evolved some postzygotic isolation.

**Reproductive isolation**
Absence or restriction of gene flow between populations over and above that due to spatial separation alone.

**Responder loci**
Loci showing deviations from Mendelian segregation (meiotic drive) due to the effect of a distorter locus.

**Secondary contact**
The meeting of the distibution ranges of two distinct populations or species after a period of evolutionary divergence in geographical isolation (allopatry).

**Sexual conflict**
The evolution of phenotypic characteristics by sexual selection, when the trait confers a fitness benefit to one sex but a fitness cost to the other.

**Sexual isolation**
Reproductive isolation as a consequence of reduced mating between members of divergent populations, including behavioural assortative mate choice and assortative fertilization in animals, as well as pollinator-mediated assortative mating in plants. Most often thought of as prezygotic, but can be postzygotic if there is disruptive sexual selection.

**Speciation continuum**
Pattern where the strength of reproductive isolation between two incipient species varies in different locations or in different species pairs that belong to the same evolutionary lineage and diverge in similar ways.

**Speciation genomics**
The field of speciation research that addresses the influence of genomic properties on the evolution of reproductive barriers and the signatures of speciation processes that are observable in genomic patterns, for example of diversity and divergence. Its aim is a conceptual and methodological integration of genomic approaches with other empirical and theoretical speciation research.

**Standing genetic variation**
Allelic variation that is currently segregating within a population; as opposed to alleles that arise through new mutation events.

**Sweep**
Increase in frequency of an allele and closely linked chromosomal segments due to positive selection. Sweeps initially reduce variation and subsequently lead to a local excess of rare alleles as new unique mutations accumulate.

**Sympatric**
Organisms, populations or species that share the same geographical region and overlap in their use of space with no spatial barriers to gene exchange.

**Transgressive phenotypes**
Expression of phenotypic variation in hybrids that exceeds the range of phenotypes observed in the parental taxa.

**Two-allele mechanism**
Reproductive barriers arise through spreading of different alleles at the same locus in two diverging populations, such as alleles for different habitat or mating preferences.

**Underdominance**
Heterozygote inferiority. The phenotype expressed in heterozygotes has lower fitness than that of either homozygote. Underdominance can be a cause of **disruptive selection**.

1. Coyne, J. & Orr, H. Speciation (Sinauer Associates, Sunderland, M.A., 2004).
2. Nosil, P. Ecological Speciation (Oxford University Press, Oxford and New York, 2012).
3. Price, T. Speciation in Birds (Roberts & Company, Greenwood Village, Colorado, 2008).
4. Dobzhansky, T. Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* **21**, 113–135 (1936).
5. Muller, H.J. & Pontecorvo, G. Recessive genes causing interspecific sterility and other disharmonies between Drosophila melanogaster and simulans. *Genetics* **27**, 157 (1942).
6. Gavrilets, S. Fitness lanscapes and the origin of species. *Princeton University Press* (2004).
7. van Doorn, G.S., Edelaar, P. & Weissing, F.J. On the Origin of Species by Natural and Sexual Selection. *Science* **326**, 1704-1707 (2009).
8. M'Gonigle, L.K., Mazzucco, R., Otto, S.P. & Dieckmann, U. Sexual selection enables long-term coexistence despite ecological equivalence. *Nature* **484**, 506-509 (2012).
9. Gavrilets, S. & Losos, J.B. Adaptive Radiation: Contrasting Theory with Data. *Science* **323**, 732-737 (2009).
10. Dieckmann, U., Doebeli, M., Metz, J.A.J. & Tautz, D. (eds.) Adaptive Speciation (Cambridge University Press, 2004).
11. Coyne, J.A. & Orr, H.A. ''Patterns of speciation in Drosophila'' revisited. *Evolution* **51**, 295-303 (1997).
12. Nolte, V., Pandey, R.V., Kofler, R. & Schlötterer, C. Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in Drosophila mauritiana. *Genome Research* **23**, 99-110 (2013).
13. van der Sluijs, I. et al. Female mating preference functions predict sexual selection against hybrids between sibling species of cichlid fish. *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 2871-2877 (2008).
14. Panhuis, T.M., Butlin, R., Zuk, M. & Tregenza, T. Sexual selection and speciation. *Trends in Ecology & Evolution* **16**, 364-371 (2001).
15. Boughman, J.W. How sensory drive can promote speciation. *Trends in Ecology & Evolution* **17**, 571-577 (2002).
16. Maan, M.E. & Seehausen, O. Ecology, sexual selection and speciation. *Ecology Letters* **14**, 591-602 (2011).
17. Hendry, A.P., Nosil, P. & Rieseberg, L.H. The speed of ecological speciation. *Functional Ecology* **21**, 455-464 (2007).
18. Abbott, R. et al. Hybridization and speciation. *Journal of Evolutionary Biology* **26**, 229-246 (2013).
19. Presgraves, D.C. Speciation Genetics: Search for the Missing Snowball. *Current Biology* **20**, R1073-R1074 (2010).
20. Presgraves, D.C. The molecular evolutionary basis of species formation. *Nature Reviews Genetics* **11**, 175-180 (2010).
21. Rieseberg, L.H. & Blackman, B.K. Speciation genes in plants. *Annals of Botany* **106**, 439-455 (2010).
22. Rieseberg, L.H. & Burke, J.M. A genic view of species integration - Commentary. *Journal of Evolutionary Biology* **14**, 883-886 (2001).
23. Bikard, D. et al. Divergent evolution of duplicate genes leads to genetic incompatibilities within A. thaliana. *Science* **323**, 623-626 (2009).
24. Masly, J.P., Jones, C.D., Noor, M.A.F., Locke, J. & Orr, H.A. Gene transposition as a cause of hybrid sterility in Drosophila. *Science* **313**, 1448-1450 (2006).
25. Ting, C.T. et al. Gene duplication and speciation in Drosophila: Evidence from the Odysseus locus. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12232-12235 (2004).
26. Matute, D.R., Butler, I.A., Turissini, D.A. & Coyne, J.A. A Test of the Snowball Theory for the Rate of Evolution of Hybrid Incompatibilities. *Science* **329**, 1518-1521 (2010).

27. Moyle, L.C. & Nakazato, T. Hybrid incompatibility “snowballs“ between Solanum species. *Science* **329**, 1521-1523 (2010).
28. Stadler, T., Florez-Rueda, A.M. & Paris, M. Testing for "Snowballing" Hybrid Incompatibilities in Solanum: Impact of Ancestral Polymorphism and Divergence Estimates. *Molecular Biology and Evolution* **29**, 31-34 (2012).
29. Lynch, M. & Force, A.G. The origin of interspecific genomic incompatibility via gene duplication. *American Naturalist* **156**, 590-605 (2000).
30. Schluter, D. Evidence for Ecological Speciation and Its Alternative. *Science* **323**, 737-741 (2009).
31. Nei, M., Maruyama, T. & Wu, C.I. Models of evolution of reproductive isolation. *Genetics* **103**, 557-579 (1983).
32. Ting, C.T., Tsaur, S.C., Wu, M.L. & Wu, C.I. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501-1504 (1998).
33. Crespi, B. & Nosil, P. Conflictual speciation: species formation via genomic conflict. *Trends in Ecology & Evolution* **28**, 48-57 (2013).
34. Frank, S.A. Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. *Evolution* **45**, 262-267 (1991).
35. Hurst, L.D. & Pomiankowski, A. Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* **128**, 841-858 (1991).
36. Cocquet, J. et al. A genetic basis for a postmeiotic X Versus Y chromosome intragenomic conflict in the mouse. *PLOS Genetics* **8**, e1002900 (2012).
37. Malik, H.S. in Progress in Molecular and Subcellular Biology (ed. Ugarovic, D.) 33-52 (Springer, 2009).
38. Campbell, P., Good, J.M. & Nachman, M.W. Meiotic sex chromosome inactivation is disrupted in sterile hybrid male house mice. *Genetics* **193**, 819-828 (2013).
39. Rebollo, R., Horard, B., Hubert, B. & Vieira, C. Jumping genes and epigenetics: Towards new species. *Gene* **454**, 1-7 (2010).
40. Watson, E.T. & Demuth, J.P. in Speciation: Natural Processes, Genetics and Biodiversity (ed. Michalak, P.) (Nova Science Publishers, 2013).
41. Burton, R.S. & Barreto, F.S. A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Molecular Ecology* **21**, 4942-4957 (2012).
42. Sambatti, J.B.M., Ortiz-Barrientos, D., Baack, E.J. & Rieseberg, L.H. Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecology Letters* **11**, 1082-1091 (2008).
43. van Doorn, G.S. & Kirkpatrick, M. Turnover of sex chromosomes induced by sexual conflict. *Nature* **449**, 909-912 (2007).
44. Charlesworth, D. & Charlesworth, B. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet. Res.* **35**, 205–214 (1980).
45. Kitano, J. et al. A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**, 1079-1083 (2009).
46. Parnell, N.F. & Streelman, J.T. Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. *Heredity* **110**, 239-246 (2013).
47. Yeaman, S. & Whitlock, M.C. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897-1911 (2011).
48. Silver, L. The peculiar journey of a selfish chromosome: mouse t-haplotypes and meiotic drive. *Trends in Genetics* **9**, 250-254 (1993).
49. Larracuente, A.M. & Presgraves, D.C. The selfish segregation distorter gene complex of Drosophila melanogaster. *Genetics* **192**, 33-53 (2012).
50. Bull, J.J. Evolution of Sex-determining Mechanisms. (Benjamin/Cumings, London, 1983).
51. Phadnis, N. & Orr, H.A. A Single Gene Causes Both Male Sterility and Segregation Distortion in Drosophila Hybrids. *Science* **323**, 376-379 (2009).

941 52. Saether, S.A. et al. Sex chromosome-linked species recognition and evolution of reproductive
942 isolation in flycatchers. *Science* **318**, 95-97 (2007).
943 53. Lindholm, A. & Breden, F. Sex chromosomes and sexual selection in poeciliid fishes. *American*
944 *Naturalist* **160**, S214-S224 (2002).
945 54. Reeve, H.K. & Pfennig, D.W. Genetic biases for showy males: Are some genetic systems
946 especially conducive to sexual selection? *Proceedings of the National Academy of Sciences of*
947 *the United States of America* **100**, 1089-1094 (2003).
948 55. Templeton, A.R. Mechanisms of Speciation - a Population Genetic Approach. *Annual Review*
949 *of Ecology and Systematics* **12**, 23-48 (1981).
950 56. Ayala, F.J. & Tracey, M.L. Genetic differentiation within and between species of Drosophila
951 willistoni-group. *Proceedings of the National Academy of Sciences of the United States of*
952 *America* **71**, 999-1003 (1974).
953 57. Coyne, J.A. Genetics and speciation. *Nature* **355**, 511-515 (1992).
954 58. Price, T.D. & Bouvier, M.M. The evolution of F-1 postzygotic incompatibilities in birds.
955 *Evolution* **56**, 2083-2089 (2002).
956 59. Stelkens, R.B., Young, K.A. & Seehausen, O. The Accumulation of Reproductive
957 Incompatibilities in African Cichlid Fish. *Evolution* **64**, 617-632 (2010).
958 60. Scopece, G., Musacchio, A., Widmer, A. & Cozzolino, S. Patterns of reproductive isolation in
959 Mediterranean deceptive orchids. *Evolution* **61**, 2623-2642 (2007).
960 61. Orr, H.A. & Turelli, M. The evolution of postzygotic isolation: Accumulating Dobzhansky-
961 Muller incompatibilities. *Evolution* **55**, 1085-1094 (2001).
962 62. Strasburg, J.L. & Rieseberg, L.H. Interpreting the estimated timing of migration events
963 between hybridizing species. *Molecular Ecology* **20**, 2353-2366 (2011).
964 63. Nosil, P., Funk, D.J. & Ortiz-Barrientos, D. Divergent selection and heterogeneous genomic
965 divergence. *Molecular Ecology* **18**, 375-402 (2009).
966 64. Strasburg, J.L. et al. What can patterns of differentiation across plant genomes tell us about
967 adaptation and speciation? *Philosophical Transactions of the Royal Society B-Biological*
968 *Sciences* **367**, 364-373 (2012).
969 65. Barton, N. Gene flow and speciation. *Heredity* **50**, 213-213 (1983).
970 66. Wu, C.I. The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-
971 865 (2001).
972 67. Via, S. & West, J. The genetic mosaic suggests a new role for hitchhiking in ecological
973 speciation. *Molecular Ecology* **17**, 4334-4345 (2008).
974 68. Feder, J.L., Egan, S.P. & Nosil, P. The genomics of speciation-with-gene-flow. *Trends in*
975 *Genetics* **28**, 342-350 (2012).
976 69. Nadeau, N.J. et al. Genomic islands of divergence in hybridizing Heliconius butterflies
977 identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society*
978 *B-Biological Sciences* **367**, 343-353 (2012).
979 70. Andrew, R.L. & Rieseberg, L.H. Divergence is focused on few genomic regions early in
980 speciation: incipient speciation of sunflower ecotypes *Evolution* **67**, 2468-2482 (2013).
981 71. Stolting, K.N. et al. Genomic scan for single nucleotide polymorphisms reveals patterns of
982 divergence and gene flow between ecologically divergent species. *Molecular Ecology* **22**, 842-
983 855 (2013).
984 72. Hohenlohe, P.A. et al. Population Genomics of Parallel Adaptation in Threespine Stickleback
985 using Sequenced RAD Tags. *Plos Genetics* **6** (2010).
986 73. Deagle, B.E. et al. Population genomics of parallel phenotypic evolution in stickleback across
987 stream-lake ecological transitions. *Proceedings of the Royal Society B-Biological Sciences* **279**,
988 1277-1286 (2012).
989 74. Jones, F.C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*
990 **484**, 55-61 (2012).
991 75. Lawniczak, M.K.N. et al. Widespread Divergence Between Incipient Anopheles gambiae
992 Species Revealed by Whole Genome Sequences. *Science* **330**, 512-514 (2010).

76. Renaut, S. et al. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* (2013).
77. Keller, I. et al. Population genomic signatures of divergent adaptation, gene flow, and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, doi: 10.1111/mec.12083 (2012).
78. Jones, F.C. et al. A Genome-wide SNP Genotyping Array Reveals Patterns of Global and Repeated Species-Pair Divergence in Sticklebacks. *Current Biology* **22**, 83-90 (2012).
79. Parchman, T.L. et al. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology* **22**, 3304-3317 (2013).
80. Gagnaire, P.-A., Pavey, S.A., Normandeau, E. & Bernatchez, L. The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish pairs assessed by RAD-sequencing. *Evolution* **67**, 2483-2497 (2013).
81. Michel, A.P. et al. Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9724-9729 (2010).
82. Via, S., Conte, G., Mason-Foley, C. & Mills, K. Localizing F-ST outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Molecular Ecology* **21**, 5546-5560 (2012).
83. Roesti, M., Hendry, A.P., Salzburger, W. & Berner, D. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology* **21**, 2852-2862 (2012).
84. Garrigan, D. et al. Genome sequencing reveals complex speciation in the Drosophila simulans clade. *Genome Research* **22**, 1499-1511 (2012).
85. Neafsey, D.E. et al. SNP Genotyping Defines Complex Gene-Flow Boundaries Among African Malaria Vector Mosquitoes. *Science* **330**, 514-517 (2010).
86. Nosil, P. et al. Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B-Biological Sciences* **279**, 5058-5065 (2012).
87. Martin, S.H. et al. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research* (2013).
88. Noor, M.A.F. & Bennett, S.M. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103**, 439-444 (2009).
89. Hahn, M.W., White, B.J., Muir, C.D. & Besansky, N.J. No evidence for biased co-transmission of speciation islands in Anopheles gambiae. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 374-384 (2012).
90. Sousa, V. & Hey, J. Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics* **14**, 404-414 (2013).
91. Scally, A. et al. Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169-175 (2012).
92. Charlesworth, B., Morgan, M.T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-1303 (1993).
93. Spencer, C.C.A. et al. The influence of recombination on human genetic diversity. *Plos Genetics* **2**, 1375-1385 (2006).
94. Nachman, M.W. & Payseur, B.A. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**, 409-421 (2012).
95. Nei, M. Molecular evolutionary genetics (Columbia University Press, New York, NY, 1987).
96. Charlesworth, B. Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**, 538-543 (1998).
97. Renaut, S., Owens, G.L. & Rieseberg, L.H. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology* (accepted).
98. Ellegren, H. et al. The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**, 756-760 (2012).

99. Consortium, H.G. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94-98 (2012).
100. Janousek, V. et al. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between Mus musculus musculus and M. m. domesticus. *Molecular Ecology* **21**, 3032-3047 (2012).
101. Elmer, K.R. & Meyer, A. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution* **26**, 298-306 (2011).
102. Nosil, P., Egan, S.P. & Funk, D.J. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**, 316-336 (2008).
103. Campbell, D. & Bernatchez, L. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution* **21**, 945-956 (2004).
104. Garrigan, D. et al. Genome sequencing reveals complex speciation in the Drosophila simulans clade. *Genome Research* **22**, 1499-1511 (2012).
105. Excoffier, L., Hofer, T. & Foll, M. Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298 (2009).
106. Gagnaire, P.-A., Normandeau, E., Pavey, S.A. & Bernatchez, L. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (Coregonus clupeaformis). *Molecular Ecology* **22**, 3036-3048 (2013).
107. Nosil, P., Parchman, T.L., Feder, J.L. & Gompert, Z. Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in Timema stick insects. *Bmc Evolutionary Biology* **12** (2012).
108. Lindtke, D., González-Martínez, S.C., Macaya-Sanz, D. & Lexer, C. Admixture mapping of quantitative traits in Populus hybrid zones: power and limitations. *Heredity* (2013).
109. Malek, T.B., Boughman, J.W., Dworkin, I. & Peichel, C.L. Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Molecular Ecology* **21**, 5265-5279 (2012).
110. Schluter, D. & Conte, G.L. Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9955-9962 (2009).
111. Seehausen, O. et al. Speciation through sensory drive in cichlid fish. *Nature* **455**, 620-626 (2008).
112. Feder, J.L. et al. Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10314-10319 (2003).
113. Nadeau, N.J. et al. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology* **22**, 814-826 (2013).
114. Seehausen, O. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**, 198-207 (2004).
115. Smadja, C.M. & Butlin, R.K. A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology* **20**, 5123-5140 (2011).
116. Bierne, N., Welch, J., Loire, E., Bonhomme, F. & David, P. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**, 2044-2072 (2011).
117. Hermann, K. et al. Tight Genetic Linkage of Prezygotic Barrier Loci Creates a Multifunctional Speciation Island in Petunia. *Current Biology* **23**, 873-877 (2013).
118. Bank, C., Burger, R. & Hermisson, J. The Limits to Parapatric Speciation: Dobzhansky-Muller Incompatibilities in a Continent-Island Model. *Genetics* **191**, 845-U345 (2012).
119. Wright, K.M., Lloyd, D., Lowry, D.B., Macnair, M.R. & Willis, J.H. Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in Mimulus guttatus. *Plos Biology* **11** (2013).
120. Barton, N.H. Multilocus clines. *Evolution* **37**, 454-471 (1983).

121. Barton, N.H. & de Cara, M.A.R. The evolution of strong reproductive isolation. *Evolution* **63**, 1171-1190 (2009).
122. Nosil, P., Harmon, L.J. & Seehausen, O. Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* **24**, 145-156 (2009).
123. Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421**, 259-264 (2003).
124. Butlin, R.K. & Ritchie, M.G. Pulling together or pulling apart: hybridization in theory and practice. *Journal of Evolutionary Biology* **26**, 294-298 (2013).
125. Servedio, M.R. & Noor, M.A.F. The role of reinforcement in speciation: Theory and data. *Annual Review of Ecology Evolution and Systematics* **34**, 339-364 (2003).
126. Qvarnström, A. & Bailey, R.I. Speciation through evolution of sex-linked genes. *Heredity* **102**, 4-15 (2009).
127. Pryke, S.R. & Griffith, S.C. Postzygotic Genetic Incompatibility between Sympatric Color Morphs. *Evolution* **63**, 793-798 (2009).
128. Pryke, S.R. Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. *Evolution* **64**, 1301-1310 (2010).
129. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419-434 (2006).
130. Felsenstein, J. Skepticism Towards Santa Rosalia, or Why Are There So Few Kinds of Animals. *Evolution* **35**, 124-138 (1981).
131. Hopkins, R. & Rausher, M.D. Pollinator-Mediated Selection on Flower Color Allele Drives Reinforcement. *Science* **335**, 1090-1092 (2012).
132. Hopkins, R. & Rausher, M.D. Identification of two genes causing reinforcement in the Texas wildflower Phlox drummondii. *Nature* **469**, 411-+ (2011).
133. Servedio, M.R., Van Doorn, G.S., Kopp, M., Frame, A.M. & Nosil, P. Magic traits in speciation: 'magic' but not rare? *Trends in Ecology & Evolution* **26**, 389-397 (2011).
134. Lowry, D.B. & Willis, J.H. A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. *Plos Biology* **8** (2010).
135. Joron, M. et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203-U102 (2011).
136. Noor, M.A.F., Grams, K.L., Bertucci, L.A. & Reiland, J. Chromosomal inversions and the reproductive isolation of species *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12084-12088 (2001).
137. Shaw, K.L. & Lesnick, S.C. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9737-9742 (2009).
138. Merrill, R.M., Van Schooten, B., Scott, J.A. & Jiggins, C.D. Pervasive genetic associations between traits causing reproductive isolation in Heliconius butterflies. *Proceedings of the Royal Society B-Biological Sciences* **278**, 511-518 (2011).
139. Saetre, G.P. & Saether, S.A. Ecology and genetics of speciation in Ficedula flycatchers. *Molecular Ecology* **19**, 1091-1106 (2010).
140. Bimova, B.V. et al. Reinforcement selection acting on the European house mouse hybrid zone. *Molecular Ecology* **20**, 2403-2424 (2011).
141. Teeter, K.C. et al. The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* **64**, 472-485 (2010).
142. Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. & Wang, J. SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. *Plos One* **7** (2012).
143. Orr, H.A. The genetics of species differences. *Trends in Ecology & Evolution* **16**, 343-350 (2001).

144. Haller, B.C., De Leon, L.F., Rolshausen, G., Gotanda, K.M. & Hendry, A.P. Magic traits: distinguishing the important from the trivial. *Trends in Ecology & Evolution* **27**, 4-5 (2012).
145. Fisher, R.A. The genetical theory of natural selection (Clarendon Press, Oxford, 1930).
146. Orr, H.A. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935-949 (1998).
147. Orr, H.A. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* **6**, 119-127 (2005).
148. Rockman, M.V. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**, 1-17 (2012).
149. Gordo, I. & Campos, P.R.A. Evolution of clonal populations approaching a fitness peak. *Biology Letters* **9**, 20120239 (2013).
150. Guillaume, F. & Whitlock, M.C. Effects of migration on the genetic covariance matrix. *Evolution* **61**, 2398-2409 (2007).
151. Bomblies, K. & Weigel, D. Arabidopsis and relatives as models for the study of genetic and genomic incompatibilities. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 1815-1823 (2010).
152. Leary, G.P. et al. Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 14081-14086 (2012).
153. Bradshaw, H.D. & Schemske, D.W. Allele substitution at a flower colour locus produces a pollinator shift in monkey flowers. *Nature* **426**, 176-178 (2003).
154. Feder, J.L. et al. Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 7990-7994 (1994).
155. Klahre, U. et al. Pollinator Choice in Petunia Depends on Two Major Genetic Loci for Floral Scent Production. *Current Biology* **21**, 730-739 (2011).
156. Dambroski, H.R. et al. The genetic basis for fruit odor discrimination in Rhagoletis flies and its significance for sympatric host shifts. *Evolution* **59**, 1953-1964 (2005).
157. Haesler, M.P. & Seehausen, O. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Proceedings of the Royal Society B-Biological Sciences* **272**, 237-245 (2005).
158. Fan, P. et al. Genetic and Neural Mechanisms that Inhibit Drosophila from Mating with Other Species. *Cell* **154**, 89-102 (2013).
159. Ballerini, E.S. et al. QTL mapping reveals the genetic architecture of loci affecting pre- and post-zygotic isolating barriers in Louisiana Iris. *Bmc Plant Biology* **12** (2012).
160. Reed, R.D. et al. optix Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. *Science* **333**, 1137-1141 (2011).
161. Barrett, R.D.H., Rogers, S.M. & Schluter, D. Environment specific pleiotropy facilitates divergence at the Ectodysplasin locus in threespine stickleback. *Evolution* **63**, 2831–2837 (2009).
162. Arnold, S.J., Burger, R., Hohenlohe, P.A., Ajie, B.C. & Jones, A.G. Understanding the Evolution and Stability of the G-Matrix. *Evolution* **62**, 2451-2461 (2008).
163. Lande, R. Quantitative genetic-analysis of multivariate evolution, applied to brain - body size allometry. *Evolution* **33**, 402-416 (1979).
164. Schluter, D. Adaptive radiation along genetic lines of least resistance. *Evolution* **50**, 1766-1774 (1996).
165. Roff, D. The evolution of the G matrix: selection or drift? *Heredity* **84**, 135-142 (2000).
166. Martin, G., Chapuis, E. & Goudet, J. Multivariate Q(st)-F-st Comparisons: A Neutrality Test for the Evolution of the G Matrix in Structured Populations. *Genetics* **180**, 2135-2149 (2008).
167. Hansen, T.F. & Houle, D. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* **21**, 1201-1219 (2008).

168. Chenoweth, S.F., Rundle, H.D. & Blows, M.W. The Contribution of Selection and Genetic Constraints to Phenotypic Divergence. *American Naturalist* **175**, 186-196 (2010).
169. Roff, D.A. & Fairbairn, D.J. A test of the hypothesis that correlational selection generates genetic correlations. *Evolution* **66**, 2953-2960 (2012).
170. Jones, A.G., Arnold, S.J. & Bürger, R. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**, 1747-1760 (2003).
171. Agrawal, A.F., Brodie, E.D. & Rieseberg, L.H. Possible consequences of genes of major effect: transient changes in the G-matrix. *Genetica* **112**, 33-43 (2001).
172. Jones, A.G., Arnold, S.J., Bürger, R., Hohenlohe, P.A. & Uyeda, J.C. The effects of stochastic and episodic movement of the optimum on the evolution of the G-matrix and the response of the trait mean to selection. *Journal of Evolutionary Biology* **25**, 2210-2231 (2012).
173. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519-U44 (2011).
174. Kopp, A. Metamodels and phylogenetic replication: a systematic approach to the evolution of developmental pathways. *Evolution* **63**, 2771-2789 (2009).
175. Eroukhmanoff, F. & Svensson, E.I. Evolution and stability of the G-matrix during the colonization of a novel environment. *Journal of Evolutionary Biology* **24**, 1363-1373 (2011).
176. Nolte, A.W. & Tautz, D. Understanding the onset of hybrid speciation. *Trends in Genetics* **26**, 54-58 (2010).
177. Selz, O.M., Lucek, K., Young, K.A. & Seehausen, O. Relaxed trait covariance in interspecific cichlid hybrids predicts morphological diversity in adaptive radiations. *Journal of Evolutionary Biology*, 10.1111/jeb.12283 (in Press).
178. Macaya-Sanz, D. et al. Genetic analysis of post-mating reproductive barriers in hybridizing European Populus species. *Heredity* **107**, 478-486 (2011).
179. Seehausen, O. Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology* **26**, 279-281 (2013).
180. Bolnick, D.I. & Fitzpatrick, B.M. Sympatric speciation: Models and empirical evidence. *Annual Review of Ecology Evolution and Systematics* **38**, 459-487 (2007).
181. Papadopulos, A.S.T. et al. A comparative analysis of the mechanisms underlying speciation on Lord Howe Island. *Journal of Evolutionary Biology* **26**, 733-745 (2013).
182. Stern, D.L. & Orgogozo, V. The loci of evolution: How predictable is genetic evolution ? *Evolution* **62**, 2155-2177 (2008).
183. Hoekstra, H.E. & Coyne, J.A. The locus of evolution: Evo devo and the genetics of adaptation. *Evolution* **61**, 995-1016 (2007).
184. Eriksson, A. & Manica, A. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13956-13960 (2012).
185. Bazin, E., Dawson, K.J. & Beaumont, M.A. Likelihood-Free Inference of Population Structure and Local Adaptation in a Bayesian Hierarchical Model. *Genetics* **185**, 587-602 (2010).
186. Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. Inference of Population Structure using Dense Haplotype Data. . *PLoS Genetics* **8**, e1002453 (2012).
187. Slate, J. Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Molecular Ecology* **14**, 363-379 (2005).
188. Buerkle, C.A. & Lexer, C. Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution* **23**, 686-694 (2008).
189. Gompert, Z. & Buerkle, C.A. Bayesian estimation of genomic clines. *Molecular Ecology* **20**, 2111-2127 (2011).
190. Trier, C.N., Hermansen, J.S., Sætre, G.P. & Bailey, R.I. Evidence for mito-nuclear and sex-linked incompatibilities between the hybrid Italian sparrow and its parent species. *PLoS Genetics*, 10.1371/journal.pgen.1004075 (in press).

191. Barrett, R.D.H., Rogers, S.M. & Schluter, D. Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255-257 (2008).
192. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nature Reviews Genetics* **10**, 184-194 (2009).
193. Orr, H.A. The Population-Genetics of Speciation - the Evolution of Hybrid Incompatibilities. *Genetics* **139**, 1805-1813 (1995).
194. Vonlanthen, P. et al. Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482**, 357-U1500 (2012).
195. Rosenblum, E.B. et al. Goldilocks Meets Santa Rosalia: An Ephemeral Speciation Model Explains Patterns of Diversification Across Time Scales. *Evolutionary Biology* **39**, 255-261 (2012).
196. Butlin, R.K., Bridle, J.R. & Schluter, D. Speciation and Patterns of Diversity (Cambridge University Press, 2009).
197. Sanderson, M.J. & Donoghue, M.J. Reconstructing shifts in diversification rates on phylogenetic trees. *Trends in Ecology & Evolution* **11**, 15-20 (1996).
198. Seehausen, O., vanAlphen, J.J.M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808-1811 (1997).
199. Taylor, E. et al. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Molecular Ecology* **15**, 343-355 (2006).
200. Seehausen, O., Takimoto, G., Roy, D. & Jokela, J. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology* **17**, 30-44 (2008).
201. Etges, W.J., de Oliveira, C.C., Noor, M.A.F. & Ritchie, M.G. Genetics of incipient speciation in Drosophila mojavensis. III. Life-history divergence in allopatry and reproductive isolation. *Evolution* **64**, 3549-3569 (2010).

**Box 1: Genomic tools for studying speciation**

Next-generation sequencing (NGS) is rapidly expanding the tool box available for studying speciation.

**Patterns of genomic divergence**: Several methods can be used to investigate genome-wide divergence along the speciation continuum. These methods include: genome scans using SNP arrays[78], RAD-seq[72, 77] or related genotyping-by-sequencing (GBS) methods, whole exome or transcriptome sequencing[76] and whole genome re-sequencing[113] of population samples.. Patterns in genome-wide divergence can be visualized and compared by means of $F_{ST}$ kernel density plots (Fig. 2) and Manhattan plots[98].

**Testing for signatures of introgression**: Various approaches are available to assess if genetic variants are shared between incipient species as a result of hybridization or due to incomplete lineage sorting[90]. The ABBA-BABA test[184] is particularly applicable to genome-scale datasets. It relies on the frequencies of two specific patterns of allele sharing among a group of four species.

**Identifying signatures of selection**: Genome scans can reveal genomic regions that show evidence of divergent selection between incipient species using **$F_{ST}$-outlier analysis** or related approaches, which can be applied to individual SNPs[77] or to smoothed average $F_{ST}$s[72] within windows or regions of the genome. The latest methods can account for demographic and other sources of variation (e.g. [105, 185]) and make improved use of high-density marker information[186].

**Mapping genes that are involved in reproductive isolation**: Genome scans of incipient species pairs along the speciation continuum are a logical first step in the search for candidate RI genes[69, 72, 74, 98]. A range of genetic mapping tools are available to identify links between divergent genomic regions and the phenotypic traits that contribute to RI. **Quantitative trait locus (QTL)** mapping is one powerful such method[187]. In short, a genome-wide set of markers is genotyped in a phenotypically variable population with known pedigree to statistically associate markers (QTLs) with phenotypes of interest (in this case traits associated with RI). With functional information on genes in the vicinity of a QTL, candidate RI genes can be identified.

**Admixture mapping**: If pedigree data are not available, it is possible to take advantage of the phenotypic and genetic differences that exist between hybridizing taxa and use admixture as the basis for genetic mapping of phenotypes that contribute to RI[109, 188] using samples from wild hybrid populations. Intrinsic and extrinsic postzygotic barriers involve alleles that are selected against in hybrids and a variety of methods can be used to identify such alleles in hybrid zones or in other situations where admixture occurs. Genomic cline analysis[189] is one such method in which candidate RI loci with low levels of introgression relative to most of the genome can be identified[79, 190].

**Manipulative selection experiments**: QTL and admixture mapping have an unfortunate bias toward detecting loci of large effect[148]. Alternatively, alleles affecting fitness and RI can be located using manipulative selection experiments which track allelic changes or genome-wide responses[86, 191]. Estimates of these effects can be ascertained by measuring selection and introgression in the wild. To date very few studies have taken this approach and none has measured effects on reproductive isolation.

**Gene expression studies**: To further investigate the significance of candidate RI-loci, expression QTL (eQTL) analysis can be useful. It identifies genomic loci that regulate expression levels of mRNAs[192]. Systematically generated eQTL information can provide insight into the mechanism underlying reproductive isolation in regions identified through genome-wide association studies, and can help to identify networks of genes and the role of gene interaction (including epistasis in DMIs) in reproductive isolation.

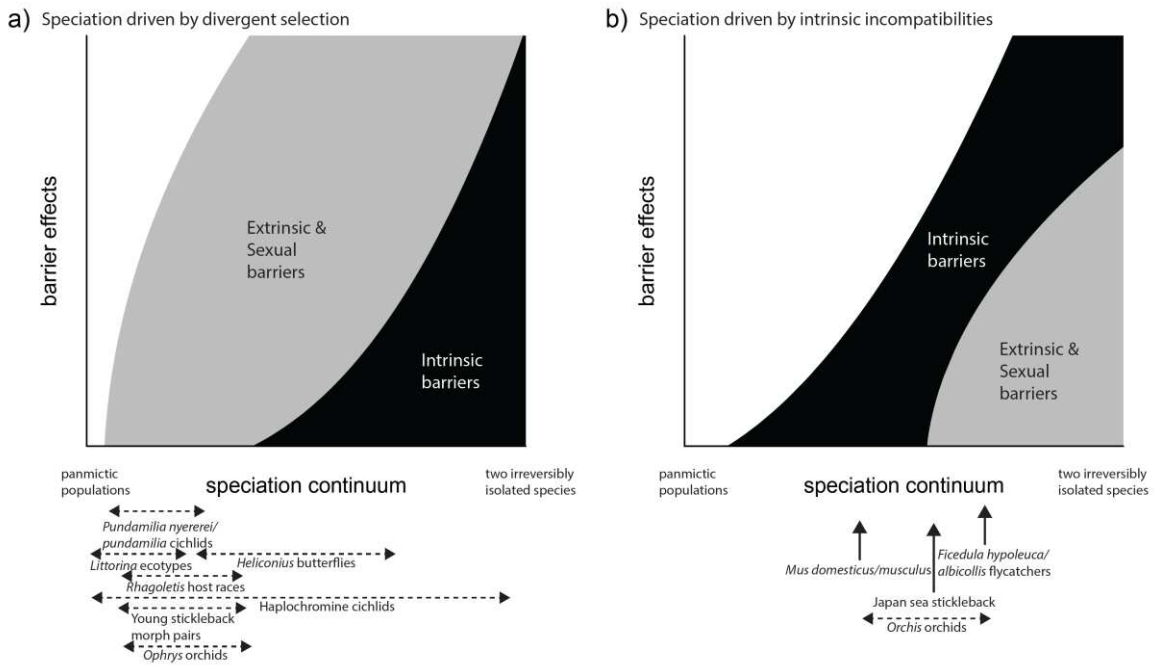
**Box 2. Evolution of reproductive isolation**

Reproductive isolation (RI) can usefully be divided into three forms: i) Extrinsic forms of postzygotic isolation result from divergent ecological or sexual selection and depend on interaction with the environment or with other individuals (e.g. reduced viability or fertility of migrants and hybrids due to ecological or behavioral factors). ii) Intrinsic forms of postzygotic isolation are due to environment independent genetic incompatibilities (e.g. Bateson-Dobzhansky-Muller incompatibilities). iii) Finally, prezygotic isolation includes phenological isolation, habitat isolation, and sexual isolation due to assortative mating or fertilization.

In speciation driven by divergent ecological or sexual selection, extrinsic and prezygotic forms of isolation evolve first, and often interact, to produce reproductive isolation, and intrinsic forms of isolation will often only evolve later in the speciation process (Panel A). In contrast, speciation driven by intrinsic barriers often results from epistatic incompatibilities, which may (though do not necessarily[19]) accumulate in an accelerating "snowball" fashion[61, 193] as a by-product of selection or due to genetic drift (the latter only slowly). Extrinsic postzygotic and prezygotic barriers may accumulate later, facilitating ecological coexistence between sibling species and reinforcement of reproductive isolation (Panel B).

In both panels the x-axis depicts the position of a diverging taxon pair on the "speciation continuum" (in terms of relative time) and the y-axis represents the strength of reproductive isolation (RI) between sister taxa. Curve shapes are hypothetical, and reflect the idea that in speciation driven by divergent selection, extrinsic postzygotic and sexual barriers arise rapidly early in speciation. Classes of barriers within each panel are not necessarily additive or interactive, and the emergence of RI via either of these barrier types should be viewed as independent trajectories. Movement along the speciation continuum, from weakly isolated species to irreversibly isolated ones, is not constant, speciation can go back and forth, or be arrested at intermittent stages, and the average timescales for speciation via the processes contrasted here (Panels A & B) may vary.

Arrows along the x-axis indicate the position(s) of model systems (studied by the authors of this paper) along the speciation continuum. These organisms vary in the strength and types of barriers isolating incipient and sister species. Studies of the genomics of speciation at different points on the speciation continuum are emerging in several systems, mainly where speciation is driven by divergent selection (as indicated by the dashed arrows showing timespans along the speciation continuum). In many cases strong reproductive isolation may never evolve, particularly in ecological speciation (e.g. [122]). Incomplete reproductive isolation may facilitate cases of "speciation reversal" (e.g.[194]) and "ephemeral" speciation (e.g. [195]).




**Box 3: New data for new theory: speciation genomics and patterns in biodiversity**

As speciation produces the raw material for biodiversity patterns, connecting speciation processes to these patterns in biodiversity is an important goal[196]. We envisage that speciation genomics can make important and unique contributions to elucidating these connections. Study of the distribution of species richness among clades provides evidence for non-uniform diversification rates among taxa, which can arise from differences in speciation and/or extinction rate (e.g.[197]). Speciation rates estimated from the fossil record are far slower than those predicted from mathematical models and observed in studies of recent diversification, and one explanation for this discrepancy is a high frequency of "ephemeral speciation," in which taxa that have recently undergone speciation have high rates of extinction[195]. This has been documented in cases of "speciation reversal"[194, 198, 199] which is possible when speciation does not reach "completion"[122, 200].

A better understanding of the genomic basis of speciation might help us to understand the influence of speciation on species persistence and patterns of species diversity. For instance, ecological speciation readily and rapidly produces divergent, partially isolated ecotypes and species that may immediately be able to coexist without competitive exclusion. Ecological speciation might thereby contribute disproportionately to the buildup of biodiversity compared to non-ecological mechanisms[196]. However, isolation between young ecologically differentiated species is often extrinsically based and contingent upon the persistence of divergent selection (see Box 2). The species that arise most rapidly may therefore be those species that are most vulnerable to extinction early in their histories[200] . In contrast, speciation via intrinsic mechanisms may produce species that are less prone to ephemerality because speciation reversal may be less likely. However, speciation rates might be slower in these lineages than in lineages where ecological speciation is common, and ecological differences must evolve after speciation in order for closely related taxa to coexist. Progress in connecting speciation to broader-scale patterns of species richness will require attention to how speciation mechanisms, and their genomic basis, influence rates of speciation and the persistence and coexistence of young species. If mechanisms of speciation leave distinctive genomic signatures, correlation between genomic patterns and disparity in species richness among clades could be tested quantitatively using comparative phylogenetic approaches.

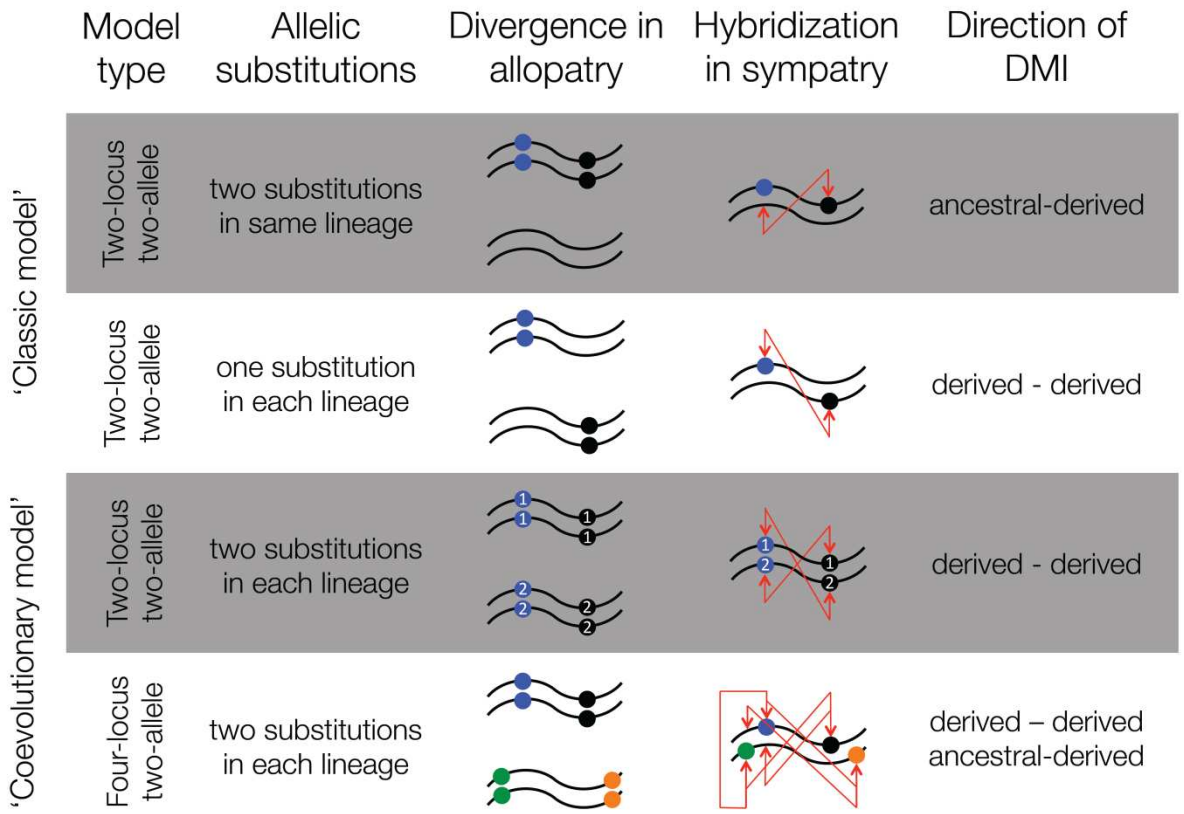



**Fig. 1. 'Classic' and coevolutionary models of hybrid incompatibility in a genomic conflict scenario**. In the 'classic model', Bateson-Dobzhansky-Muller incompatibilities (DMIs) are envisioned as two-locus, two-allele interactions, in which incompatibilities arise between an ancestral allele and an allele derived in one lineage (1st row) or between alleles derived in two separate lineages (2nd row); a special case of the latter model can refer to maternal-effect selfish loci in which maternal "poison" and zygotic "antidote" are due to developmental expression divergence of the same locus. In the coevolutionary models, DMIs are continually fixed at the same loci (3rd row) or at different loci (4th row). In all examples with two substitutions in a lineage, the selfish locus (left) drives the evolution of the restorer locus (right). Red arrows indicate negative epistatic interactions between complimentary loci. In all models, the ancestral state is wild-type except for row three. In this row, the ancestral state is a coevolving selfish element-restorer system. Insight into the role of genomic conflict in speciation reveals the potential for further development of models of hybrid incompatibility. Models that incorporate the possibility for increased lag-load due to ongoing coevolution predict successively more severe incompatibilities. Additional theoretical work is needed to investigate such coevolutionary models.

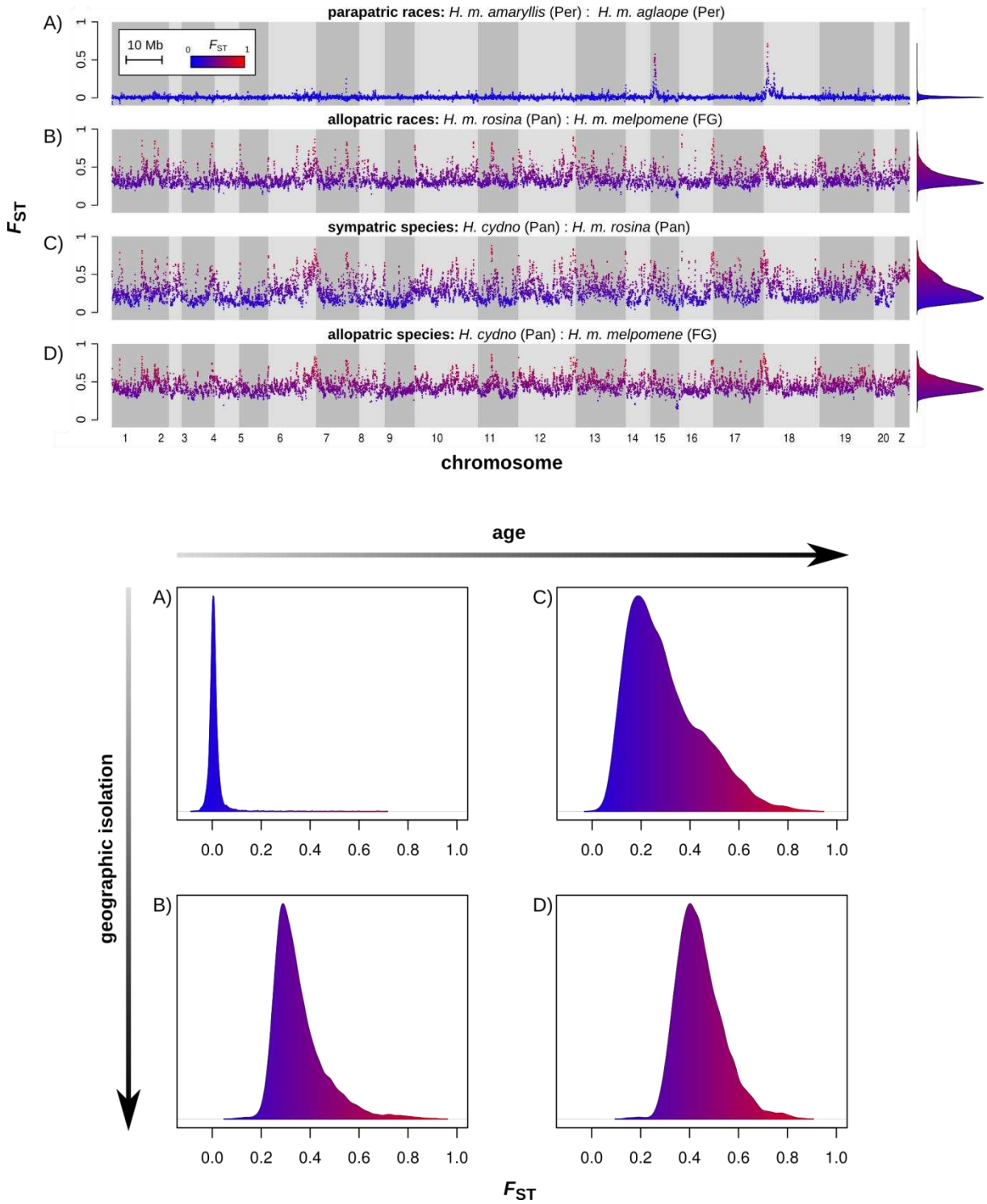
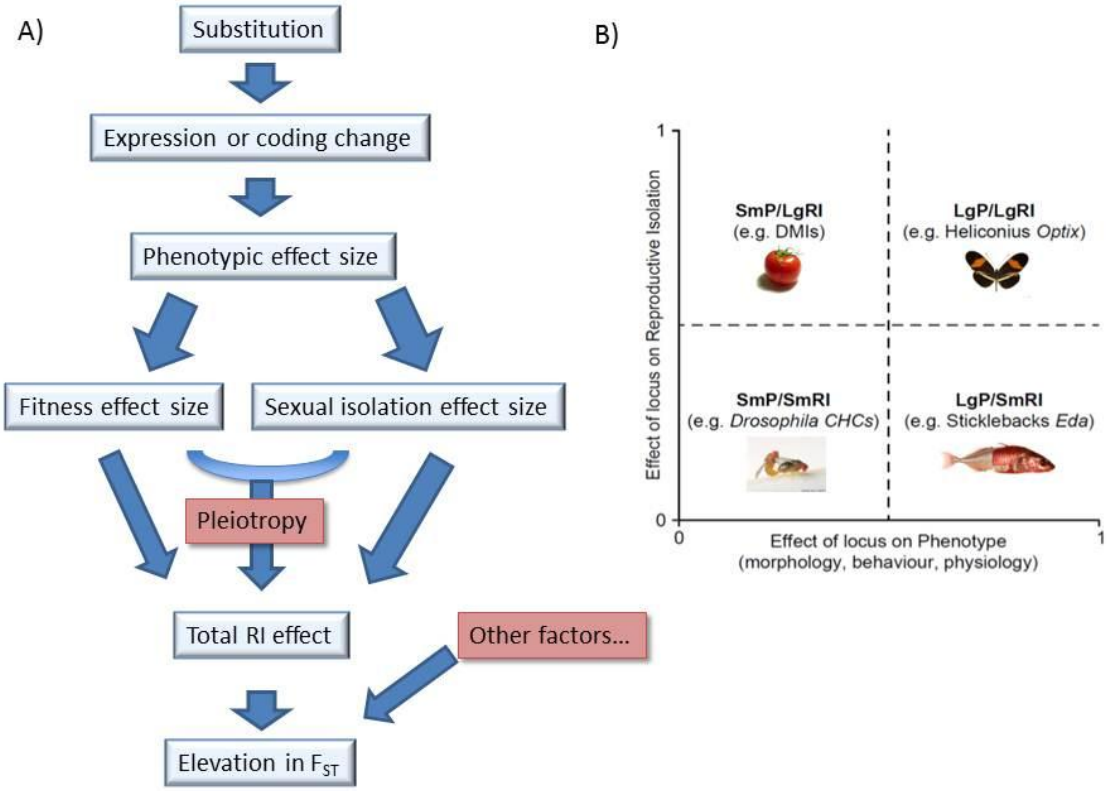
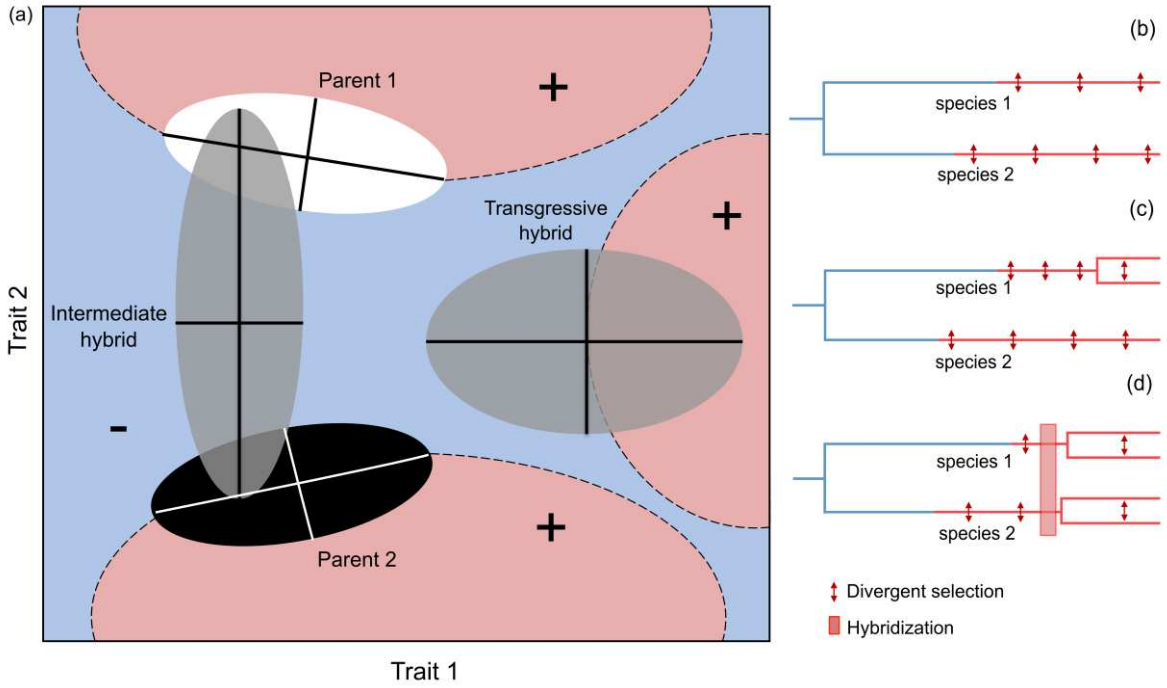



**Fig. 2**. **Genomic patterns of divergence along the speciation continuum in *Heliconius* butterflies**. The top panel shows the patterns of differentiation between hybridizing parapatric races (A) and sympatric species (C) and between geographically isolated races (B) and species (D) along the genome. Divergence is highly heterogeneous even between allopatric populations of the same species (B). The shape of the frequency distribution of locus-specific $F_{ST}$ values (bottom panel) clearly differs between the different stages in the continuum and between geographic scenarios with, for example, the greater variance in (C) consistent with gene flow between species in sympatry. However, the challenge is to distinguish between speciation with (A, C) versus without (B, D) gene flow.




1400

1401 **Fig. 3. Effect sizes of substitutions on phenotype and on reproductive isolation.** (A) Effects of variation
1402 at different levels, and connections between those levels. The size of effect can vary at each step from zero or
1403 quite small to very large. A substitution can alter gene expression or protein coding, which in turn has some
1404 effect on a phenotype. This phenotype can have effects of varying size on environment-dependent fitness (and
1405 hence possibly extrinsic postzygotic isolation), environment-independent fitness (hence possibly intrinsic
1406 postzygotic isolation) and on prezygotic isolation. Alternatively a phenotype may pleiotropically affect both
1407 fitness and prezygotic isolation. All these effects combine to generate total RI, which will likely elevate $F_{ST}$,
1408 although other factors can alter $F_{ST}$ as well. (B) The lack of correlation between the effect of a locus on
1409 phenotype (P) and on reproductive isolation (RI). An example for each of the four relationships is shown to
1410 illustrate that phenotypic effect size does not necessarily predict RI effect size: loci with small effect on
1411 phenotype and large effect on reproductive isolation (SmP/LgRI: DMIs in *Solanum*[27]); loci with large effect on
1412 phenotype and large effect on reproductive isolation (LgP/LgRI: *Optix* in *Heliconius* [160]); loci with small effect on
1413 phenotype and small effect on reproductive isolation (SmP/SmRI: CHCs in *Drosophila*[201]); loci with large effect
1414 on phenotype and small effect on reproductive isolation (LgP/SmRI: *Eda* in stickleback[191]). The relationships
1415 between phenotypic and RI effect size and $F_{ST}$ are largely unknown at present.




1416

1417

1418 **Fig. 4. Influence of genetic constraints on speciation.** (A) With the help of NGS, it is now feasible to infer
1419 relatedness of individuals in any given natural population and thus to estimate a **G**-matrix without the use of
1420 pedigree-data[173]. The **G**-matrix (represented here as an ellipse in the space of two quantitative traits) can bias
1421 evolution in certain directions, and depending on the adaptive landscape (represented by regions of higher (+;
1422 red) and lower (-, blue) fitness than the parental populations (white, black)), might constrain adaptive
1423 divergence and speciation. Hybridization events may facilitate speciation by aligning the **G**-matrix in the
1424 direction of divergence between parental species (intermediate hybrid), or by giving rise to novel phenotypes
1425 (transgressive hybrid) in new regions of positive fitness that cannot be reached through gradual evolution in
1426 either of the parental species.
1427
1428 (B-D) The influence of genetic constraints on speciation can be tested at the phylogenetic level. (B) Constraints
1429 may persist over evolutionary time as a result of the inability of divergent selection to change genetic
1430 architecture, preventing speciation from happening. (C) Alternatively, other forms of selection (e.g.
1431 correlational selection) can alter the structure and orientation of the **G**-matrix and potentially facilitate
1432 divergence and speciation over moderate time scales. (D) Hybridization and gene flow can dramatically alter **G**
1433 in just a few generations, fueling adaptive divergence and resulting in sudden bursts of speciation. Note that
1434 hybridization between sister species is shown here for illustration, but hybridization that facilitates divergence
1435 may occur more widely among related taxa.

1436

**Biographies**

Ole Seehausen studied speciation and hybridization since his PhD at the University of Leiden in the 1990s. Adaptive radiations receive his particular attention, such as the cichlid fishes of Lake Victoria and, more recently, the whitefish of prealpine European lakes, stickleback and trout. He is a professor in the Institute of Ecology & Evolution of the University of Bern and head of a research department at EAWAG, the Swiss Federal Institute of Aquatic Science and Technology. His lab combines ecological and behavioral research with genetics and genomics to investigate processes and mechanisms implicated in adaptation, speciation, species coexistence and extinction.

Roger Butlin has studied speciation since his postdoctoral work with Godfrey Hewitt in the 1980s. He is interested in the processes generating reproductive isolation and its genetic basis. Reinforcement has been a particular focus of study. Current projects are examining the role of chemosensory genes in aphid host race formation and the genetic basis of parallel local adaptation and speciation in periwinkles. He is a professor of evolutionary biology at the University of Sheffield in the UK and currently holds the 2013 Tage Erlander guest professorship at the University of Gothenburg in Sweden.

Irene Keller is a bioinformatician at the Department of Clinical Research of the University of Bern (Switzerland). She received her PhD from the University of Bern and worked as a postdoctoral fellow with Richard Nichols at Queen Mary University of London and with Jukka Jokela and Ole Seehausen at Eawag and University of Bern (Switzerland). Her interests focus on the application of molecular and bioinformatics tools to understand the genetic basis of adaptation, speciation and human disease.

Catherine E.Wagner is an evolutionary biologist with interests in speciation and the origins of diversity, and the relationships between diversity-generating processes and macroevolutionary patterns. Her research uses population genetic, phylogenetic, and comparative methods to study diversification. She is currently a postdoctoral researcher at Eawag, the Swiss Federal Institute of Aquatic Science and Technology and the University of Bern, Switzerland, where her work focuses primarily on African cichlid fishes. She earned her Ph.D. in ecology and evolutionary biology from Cornell University in 2011.

Janette Boughman and her lab study the selective forces causing speciation in threespine sticklebacks, with particular focus on sexual selection and its interaction with natural selection to generate reproductive isolation. She uses lab and field behavioral experiments to understand the subtle yet powerful action of these forces on phenotypic and genetic evolution and how this transmits to the genome. She has studied both the accumulation of reproductive isolation and its loss through reverse speciation. Recent work investigates fitness landscapes at both the phenotypic and genetic level and their role in diversification. She is Associate Professor at Michigan State University.

Paul A. Hohenlohe is an Assistant Professor in the Department of Biological Sciences and the Institute for Bioinformatics and Evolutionary Studies at the University of Idaho. He earned his Ph.D. in zoology at the University of Washington in 2000 and subsequently worked as a conservation biologist and postdoctoral researcher. His research focus is on evolutionary genetics and genomics, including RAD sequencing and other tools for population genomics and conservation in non-model organisms, experimental evolution, and evolutionary quantitative genetics theory.

Catherine Peichel earned her Ph.D. in the area of developmental genetics at Princeton University in 1998. During this time, she became intrigued by the genetic basis of phenotypic differences between species. Thus, during her postdoctoral fellowship with David Kingsley at Stanford University, she helped to develop the threespine stickleback as a genetic and genomic model system. She has led a research laboratory at the Fred Hutchinson Cancer Research Center in Seattle, Washington since 2003. Her lab takes a number of approaches to investigate the genetic and genomic changes that underlie adaptation and speciation in sticklebacks.

Glenn-Peter Sætre is a professor in evolutionary biology at the University of Oslo, Norway. He obtained his doctorate also in Oslo and worked several years at Uppsala University, Sweden as a post doc and assistant professor before returning to Oslo in 2003 as full professor. He studies speciation, hybridization and adaptive evolution, mainly in birds, combining genomic analysis and population genetics with behavioural and ecological studies. His research lab is currently mainly focusing on the genomics of hybrid speciation.

## Highlighted references


Andrew, R.L. & Rieseberg, L.H. Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes *Evolution* **67**, 2468-2482 (2013).
*Applying genomic analysis at different points on the speciation continuum is important for understanding how reproductive isolation develops. This study shows how differentiation can be focused on a small proportion of the genome early in speciation.*

Barrett, R. D. H., Rogers, S. M. & Schluter, D. Natural selection on a major armor gene in threespine stickleback. Science 322, 255-257 (2008).
*This study showed unexpected complexity in the response to selection for alleles of the Ectodysplasin locus, probably because of pleiotropic effects on other fitness-related traits. Such pleiotropic effects may be widespread and have major impacts on the progress of adaptation and speciation.*

Barton, N. H. & de Cara, M. A. R. The evolution of strong reproductive isolation. Evolution 63, 1171-1190 (2009).
*An important, theoretical paper considering the conditions under which selection can overcome recombination to bring together multiple reproductive barriers and so generate strong reproductive isolation.*

Bimova, B. V. et al. Reinforcement selection acting on the European house mouse hybrid zone. Mol. Ecol. 20, 2403-2424 (2011).
*Using powerful analyses of the exceptionally well-studied mouse hybrid zone, the authors provide clear evidence for the operation of reinforcement and also for the limits on its effectiveness in reducing gene flow.*

Coyne, J. & Orr, H. Speciation. (Sinauer Associates, Sunderland, M.A., 2004).
*Coyne and Orr's comprehensive book is a must-read for every student of speciation. It provides excellent background information and review of all facets of research on the speciation process.*

Coyne, J. A. & Orr, H. A. ''Patterns of speciation in Drosophila'' revisited. Evolution 51, 295-303 (1997).
*A key comparative study on the rate of evolution of reproductive isolation that was a model for similar studies in other taxa*

Ellegren, H. et al. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**, 756-760 (2012).
*A truly genome-wide view of differentiation in a system with outstanding ecological and behavioural information available. Strikingly heterogeneous levels of divergence were observed, including major divergence peaks at telomeres, that are, as yet, unexplained.*

Felsenstein, J. Skepticism Towards Santa Rosalia, or Why Are There So Few Kinds of Animals. Evolution 35, 124-138 (1981).
*A key paper introducing the antagonism between recombination and the build-up of linkage disequilibrium which lies at the heart of the speciation process.*

Feder, J. L., Egan, S. P. & Nosil, P. The genomics of speciation-with-gene-flow. Trends in Genetics 28, 342-350 (2012).
*In this review, the authors bring together empirical studies with theory on the impact of divergent selection on gene flow elsewhere in the genome to examine how reproductive isolation might spread through the genome as speciation proceeds.*

Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487, 94-98 (2012).
*An excellent illustration of how annotated reference genomes together with population-level NGS data can address fundamental questions about the genetic basis for speciation. The Heliconius Genome Consortium provide evidence that genes for mimicry in color pattern have been exchanged via hybridization among Heliconius species, thus implicating adaptive introgression as a crucial component of speciation in this system.*

Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484, 55-61 (2012).
*An excellent example of how genomic data can inform researchers about the genetic basis for repeated adaptation to similar environments. They show that stickleback fish populations throughout the northern hemisphere have repeatedly utilized the same genetic elements to adapt to freshwater environments.*

Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. Genetics 173, 419-434 (2006).
*This paper gives a thorough theoretical background and new insights into the role of chromosome inversions in adaptation and speciation.*

Lindtke, D., González-Martínez, S. C., Macaya-Sanz, D. & Lexer, C. Admixture mapping of quantitative traits in *Populus* hybrid zones: power and limitations. Heredity, doi:10.1038/hdy.2013.69 (2013).
*One of the first studies to use admixture mapping in a natural hybrid population to examine the genetic basis of traits contributing to reproductive isolation. Application of this powerful approach is likely to make important contributions to speciation research.*

Nolte, V., Pandey, R.V., Kofler, R. & Schlötterer, C. Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in *Drosophila mauritiana*. *Genome Research* **23**, 99-110 (2013).
*Genome-sequence data here show how widespread strong selection due to genomic conflict can be, suggesting that it may be a potent source of incompatibilities between previously isolated populations.*

Nosil, P. Ecological Speciation (Oxford University Press, Oxford and New York, 2012).
*An in-depth treatment of speciation by divergent natural selection.*

Orr, H. A. & Turelli, M. The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. Evolution 55, 1085-1094, (2001).
*Orr and Turelli's paper provides excellent background on the evolution of reproductive isolation via postzygotic intrinsic barriers. It considers the dynamics of accumulation of Dobzhansky-Muller incompatibilities in diverging lineages.*

Presgraves, D. C. The molecular evolutionary basis of species formation. Nature Reviews Genetics 11, 175-180, (2010).
*Presgraves' review summarizes empirical evidence on the genetic basis of hybrid dysfunction, focusing on work in genetic model organisms where key genes have been identified.*

Schluter, D. Adaptive radiation along genetic lines of least resistance. Evolution 50, 1766-1774 (1996).
*A key conceptual paper showing how the structure of genetic variances and covariances among quantitative traits can influence the direction of evolution and so the progress of adaptive radiation.*

Seehausen, O. et al. Speciation through sensory drive in cichlid fish. Nature 455, 620-626 (2008).
*This study shows how divergent female preferences, initially with an ecological basis, can impose selection on male signal traits and generate reproductive isolation in the face of gene flow, given the right balance between intensity of selection and distribution of habitats.*

Wu, C. I. The genic view of the process of speciation. Journal of Evolutionary Biology 14, 851-865 (2001).
*Wu's thought-provoking review discusses the process of speciation from a genetic perspective, by highlighting the difference between thinking of the evolution of reproductive isolation as a whole-genome process, versus understanding the influence of specific loci on reproductive isolation/gene exchange. Wu's point that genes, and not whole genomes, are the unit of species differentiation is a seminal perspective, critical to much of the current work in speciation genetics.*

**Online key points:**

- Speciation is a central process in evolution that is fundamentally about the origin of reproductive isolation. The latest generation of genomic approaches provides remarkable opportunities to describe speciation and learn about speciation mechanisms.
- Genome scans, now truly genome-wide and at base-pair resolution, reveal substantial genomic divergence among incipient species even in the face of gene flow, with extensive genomic heterogeneity in the extent of differentiation, especially at early stages of speciation, both in sympatry and in allopatry.
- The sources of this heterogeneity remain incompletely understood. Combining genome scans with sophisticated population genetic modeling, QTL, and admixture analysis has the potential to isolate the influence of selection from demographic, historical and structural effects and to link these sources of genomic divergence to phenotypes and to reproductive isolation.
- Available empirical data suggest that differentiation between parapatric populations can be restricted to few genomic islands, whereas incipient species that coexist in sympatry show differentiation widely distributed across the genome. This may suggest that genomically widespread selection is required to permit the maintenance and perhaps the buildup of genetic differentiation in sympatry.
- Recent genomic studies reveal that the genetic basis of reproductive isolation is often complex. The effects of pleiotropy, genetic correlations, and patterns of recombination need to be considered, alongside effects of ecological and sexual selection as well as genomic conflict.
- A surprising recent discovery has been the re-use of ancient gene variants in speciation, acquired from standing genetic variation or by introgressive hybridization.
- We propose a roadmap for the development of speciation genomics towards answering classical as well as emerging questions in speciation research.